Access DB# 49545

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: _____ Examiner # :_____ Date: _____
Art Unit: _____ Phone Number 30_____ Serial Number: _____
Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

****

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

_____

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

****

STAFF USE ONLY

Searcher: G Vezza. Kelly
Searcher Phone #: 308-4501
Searcher Location: Biotech Lib.
Date Searcher Picked Up: 8/24/01
Date Completed: 8/27/01
Searcher Prep & Review Time: _____
Clerical Prep Time: 4 mi
Online Time: 2 mi

Type of Search

NA Sequence (#) 2
AA Sequence (#) _____
Structure (#) _____
Bibliographic _____
Litigation _____
Fulltext _____
Patent Family _____
Other _____

Vendors and cost where applicable

STN _____
Dialog _____
Questel/Orbit _____
Dr.Link _____
Lexis/Nexis _____
Sequence Systems ABSSOL
WWW/Internet _____
Other (specify) _____

PTO-1590 (1-2000)

STIC-Biotech/ChemLib

49545

From: Jiang, Dong
Sent: Thursday, August 23, 2001 4:34 PM
To: STIC-Biotech/ChemLib
Subject: SN09/333,159

6/9
McCarthy, S.A.,
Fraser

Please search SEQ ID NO: 45 and 46

-issued
-commercial

Please send results on paper to Dong Jiang in 10B-01 (mail stop CM1-10C01). Thank you very much.

Dong Jiang (78243)
703-305-1345
U.S. Patent and Trademark Office
Art Unit 1646
dong.jiang@uspto.gov
CM1-10B01
Mail stop: CM1-10C01

Try to find the art on ≥ 54 nt (for 18 aa.)
60% var.

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: August 24, 2001, 22:38:56 ; Search time 97.86 Seconds
(without alignments)
2454.893 Million cell updates/sec

Title: US-09-333-159-46
Perfect score: 1269
Sequence: 1 atgttggaaaccttgtcaag.....gacggtgtgaggccgtattg 1269

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	444.2	35.0	1137	1	US-08-227-108-2	Sequence 2, Appli
	2	444.2	35.0	1137	2	US-09-073-674-2	Sequence 2, Appli
	3	444.2	35.0	1140	1	US-08-227-108-4	Sequence 4, Appli
	4	444.2	35.0	1140	2	US-09-073-674-4	Sequence 4, Appli
	5	444.2	35.0	1146	1	US-08-227-108-6	Sequence 6, Appli
	6	444.2	35.0	1146	2	US-09-073-674-6	Sequence 6, Appli
	7	444.2	35.0	1528	1	US-08-227-108-1	Sequence 1, Appli
	8	444.2	35.0	1528	2	US-09-073-674-1	Sequence 1, Appli
	9	48.2	3.8	178	1	US-08-751-782-2	Sequence 2, Appli
	10	48.2	3.8	178	2	US-08-925-171-2	Sequence 2, Appli
	11	46.4	3.7	43676	3	US-09-356-952-12	Sequence 12, Appl
	12	36.6	2.9	3271	2	US-08-852-806-1	Sequence 1, Appli
	13	36.6	2.9	3271	3	US-09-163-669-1	Sequence 1, Appli
	14	36.6	2.9	3282	1	US-08-276-852-154	Sequence 154, App
c	15	36.6	2.9	3282	1	US-08-276-852-169	Sequence 169, App
	16	36.6	2.9	3282	1	US-08-899-575-154	Sequence 154, App
c	17	36.6	2.9	3282	1	US-08-899-575-169	Sequence 169, App
	18	36.6	2.9	3282	1	US-08-899-575-154	Sequence 154, App
c	19	36.6	2.9	3282	1	US-08-899-575-169	Sequence 169, App
	20	36.6	2.9	3282	5	PCT-US95-08743-154	Sequence 154, App
c	21	36.6	2.9	3282	5	PCT-US95-08743-169	Sequence 169, App
	22	36.6	2.9	13254	1	US-08-276-852-156	Sequence 156, App
c	23	36.6	2.9	13254	1	US-08-276-852-170	Sequence 170, App
	24	36.6	2.9	13254	1	US-08-899-575-156	Sequence 156, App
c	25	36.6	2.9	13254	1	US-08-899-575-170	Sequence 170, App
	26	36.6	2.9	13254	1	US-08-899-575-156	Sequence 156, App
c	27	36.6	2.9	13254	1	US-08-899-575-170	Sequence 170, App
	28	36.6	2.9	13254	5	PCT-US95-08743-156	Sequence 156, App
c	29	36.6	2.9	13254	5	PCT-US95-08743-170	Sequence 170, App
	30	33.6	2.6	8392	1	US-08-080-255-6	Sequence 6, Appli
	31	33.6	2.6	8392	3	US-08-465-713-6	Sequence 6, Appli
	32	33.6	2.6	8392	5	PCT-US93-05857-6	Sequence 6, Appli
c	33	32.6	2.6	7218	1	US-08-232-463-14	Sequence 14, Appl
c	34	32	2.5	2197	1	US-08-233-005-3	Sequence 3, Appli
c	35	32	2.5	2197	1	US-08-428-943-3	Sequence 3, Appli
c	36	32	2.5	2197	3	US-09-016-649-3	Sequence 3, Appli
c	37	32	2.5	2197	5	PCT-US95-04858-3	Sequence 3, Appli
	38	32	2.5	8342	3	US-08-545-860D-63	Sequence 63, Appl
	39	32	2.5	8342	5	PCT-US94-04496-63	Sequence 63, Appl
	40	31	2.4	3212	3	US-08-673-814-1	Sequence 1, Appli
c	41	30.4	2.4	1529	1	US-08-336-778-1	Sequence 1, Appli
c	42	30.2	2.4	4982	3	US-08-699-103B-1	Sequence 1, Appli
	43	30.2	2.4	10079	2	US-08-476-866-20	Sequence 20, Appl
c	44	30	2.4	3639	2	US-08-737-524B-26	Sequence 26, Appl
	45	30	2.4	5178	2	US-08-474-169-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-227-108-2
; Sequence 2, Application US/08227108
; Patent No. 5807726
; GENERAL INFORMATION:
; APPLICANT: Blanchard, Claire
; APPLICANT: Benicourt, Claude
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: Recombinant Dog Gastric Lipase
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/227,108
; FILING DATE: 03-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fanucci, Allan A.
; REGISTRATION NUMBER: 30,256
; REFERENCE/DOCKET NUMBER: 7620-033
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1137 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1137
US-08-227-108-2

Query Match 35.0%; Score 444.2; DB 1; Length 1137;
Best Local Similarity 62.8%; Pred. No. 5.6e-139;
Matches 689; Conservative 0; Mismatches 408; Indels 0; Gaps 0;

```
Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
Db     26 ACCCTGAAGTGACCATGAATATAAGTCAGATGATCACCTACTGGGGATACCCAGCTGAGG 85

Qy    185 aatatgaagtcgcaactgaagatgggtatatccctttctgttaacaggattcctcgaggcc 244
Db     86 AATATGAAGTTGTGACCGAAGACGGTTATATCCTTGGGATCGACAGAATTCCTTATGGGA 145

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
Db    146 GGAAAAATTCAGAGAATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCG 205

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
Db    206 CATCAGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCCTGGCCG 265

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
Db    266 ACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACCTGGGCCAGGAGGAATC 325

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
Db    326 TGTACTACTCGCCCGACTCCGTCGAATTCTGGGCTTTCAGCTTTGACGAGATGGCTAAAT 385

Qy    485 ttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 544
Db    386 ATGACCTTCCCGCCACCATTGACTTCATCTTGAAGAAAACGGGACAGGACAAGCTACACT 445

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
Db    446 ACGTTGGCCATTCCCAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGC 505

Qy    605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
Db    506 TGGCGAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGTACACCG 565

Qy    665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
Db    566 AAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTCAAGCTTATATTTGGAA 625

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccttttgtggcc 784
Db    626 ACAAAATATTCTACCCACACCACTTCTTTGATCAATTTCTCGCCACCGAGGTATGCTCCC 685

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
Db    686 GCGAGACGGTGGATCTCCTCTGCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTA 745

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
Db    746 TGAACTTGAACATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGG 805

Qy    905 tgcaaaatattctacactggagccaggcagtgaatttctggtgaactccgggcatttgact 964
Db    806 TTCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAGCTTTTGACT 865

Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
Db    866 GGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGCATGCCTCCCTACTACAACC 925

Qy   1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
Db    926 TGACAGACATGCATGTGCCAATCGCAGTGTGGAACGGTGGCAACGACTTGCTGGCCGACC 985

Qy   1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
Db    986 CTCACGATGTTGACCTTTTGCTTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTC 1045

Qy   1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
Db   1046 CTCCTTACAATCACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATG 1105

Qy   1205 aaatcatccatctgatg 1221
Db   1106 AAATTGTTTCCATGATG 1122
```

```
RESULT   2
US-09-073-674-2
; Sequence 2, Application US/09073674
; Patent No. 5998189
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Warner-Lambert Company
;      STREET:  2800 Plymouth Road
;      CITY:  Ann Arbor
;      STATE:  Michigan
;      COUNTRY:  U.S.A.
;      ZIP:  48105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/073,674
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crissey, Todd M.
;      REGISTRATION NUMBER:  37,807
;      REFERENCE/DOCKET NUMBER:  5072-D1-66-TMC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  734 622-7530
;      TELEFAX:  734 622-1553
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1137 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1137
US-09-073-674-2
```

```
  Query Match           35.0%;  Score 444.2;  DB 2;  Length 1137;
  Best Local Similarity  62.8%;  Pred. No. 5.6e-139;
  Matches  689;  Conservative    0;  Mismatches  408;  Indels    0;  Gaps    0;

Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
Db     26 ACCCTGAAGTGACCATGAATATAAGTCAGATGATCACCTACTGGGGATACCCAGCTGAGG 85

Qy    185 aatatgaagtcgcaactgaagatgggtatatccctttctgttaacaggattcctcgaggcc 244
Db     86 AATATGAAGTTGTGACCGAAGACGGTTATATCCTTGGGATCGACAGAATTCCTTATGGGA 145

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
Db    146 GGAAAAATTCAGAGAATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCG 205

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
Db    206 CATCAGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCCTGGCCG 265

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
```

```
Db    266 ACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACCTGGGCCAGGAGGAATC 325

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
Db    326 TGTACTACTCGCCCGACTCCGTCGAATTCTGGGCTTTCAGCTTTGACGAGATGGCTAAAT 385

Qy    485 ttgaccttcctgcagtgataaactttatttttgcagaaaacgggccaggaaaagatctatt 544
Db    386 ATGACCTTCCCGCCACCATTGACTTCATCTTGAAGAAAACGGGACAGGACAAGCTACACT 445

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
Db    446 ACGTTGGCCATTCCCAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGC 505

Qy    605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
Db    506 TGGCGAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGTACACCG 565

Qy    665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
Db    566 AAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTCAAGCTTATATTTGGAA 625

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
Db    626 ACAAAATATTCTACCCACACCACTTCTTTGATCAATTTCTCGCCACCGAGGTATGCTCCC 685

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
Db    686 GCGAGACGGTGGATCTCCTCTGCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTA 745

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
Db    746 TGAACTTGAACATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGG 805

Qy    905 tgcaaaatattctacactggagccaggcagtgaatttctggtgaactccgggcatttgact 964
Db    806 TTCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAGCTTTTGACT 865

Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
Db    866 GGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGCATGCCTCCCTACTACAACC 925

Qy   1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
Db    926 TGACAGACATGCATGTGCCAATCGCAGTGTGGAACGGTGGCAACGACTTGCTGGCCGACC 985

Qy   1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
Db    986 CTCACGATGTTGACCTTTTGCTTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTC 1045

Qy   1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
Db   1046 CTCCTTACAATCACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATG 1105

Qy   1205 aaatcatccatctgatg 1221
Db   1106 AAATTGTTTCCATGATG 1122


RESULT   3
US-08-227-108-4
; Sequence 4, Application US/08227108
; Patent No. 5807726
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/227,108
;      FILING DATE:  03-APR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fanucci, Allan A.
;      REGISTRATION NUMBER:  30,256
;      REFERENCE/DOCKET NUMBER:  7620-033
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 790-9090
;      TELEFAX:  212 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1140 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1140
US-08-227-108-4


  Query Match             35.0%;  Score 444.2;  DB 1;  Length 1140;
  Best Local Similarity   62.8%;  Pred. No. 5.6e-139;
  Matches  689;  Conservative    0;  Mismatches  408;  Indels    0;  Gaps    0;

Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
Db     29 ACCCTGAAGTGACCATGAATATAAGTCAGATGATCACCTACTGGGGATACCCAGCTGAGG 88

Qy    185 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 244
Db     89 AATATGAAGTTGTGACCGAAGACGGTTATATCCTTGGGATCGACAGAATTCCTTATGGGA 148

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
Db    149 GGAAAAATTCAGAGAATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCG 208

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
Db    209 CATCAGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCCTGGCCG 268

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
Db    269 ACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACCTGGGCCAGGAGGAATC 328

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
Db    329 TGTACTACTCGCCCGACTCCGTCGAATTCTGGGCTTTCAGCTTTGACGAGATGGCTAAAT 388

Qy    485 ttgaccttcctgcagtgataaactttatttttgcagaaaacgggccaggaaaagatctatt 544
Db    389 ATGACCTTCCCGCCACCATTGACTTCATCTTGAAGAAAACGGGACAGGACAAGCTACACT 448

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
Db    449 ACGTTGGCCATTCCCAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGC 508

Qy    605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
Db    509 TGGCGAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGTACACCG 568
```

```
Qy    665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
          ||| ||      | |||| |  || |  | ||     |  |  |||||    | ||||| |
Db    569 AAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTCAAGCTTATATTTGGAA 628

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
          | |||  |||       ||   |   |||     |   || |  |   |  | ||   ||
Db    629 ACAAAATATTCTACCCACACCACTTCTTTGATCAATTTCTCGCCACCGAGGTATGCTCCC 688

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
            | ||   | ||||   | || || ||   | ||||  |  |  |||||||  |||| |
Db    689 GCGAGACGGTGGATCTCCTCTGCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTA 748

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
            ||  |||||||||| ||      ||| |||    | || | ||  || ||||||| |
Db    749 TGAACTTGAACATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGG 808

Qy    905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
          | || ||  | || ||||||  |||||| || || |||||  |  |||  || |||||||
Db    809 TTCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAGCTTTTGACT 868

Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
          |||| ||        | ||  ||    | |   |||||   |  |||      ||||
Db    869 GGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGCATGCCTCCCTACTACAACC 928

Qy   1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
          | | ||| |||   || || |  ||| ||||||  || ||  | |||| |||  |  | |
Db    929 TGACAGACATGCATGTGCCAATCGCAGTGTGGAACGGTGGCAACGACTTGCTGGCCGACC 988

Qy   1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
          |  | || ||  |  |  |||| ||  || |  |||| ||||  ||||  | ||| ||||
Db    989 CTCACGATGTTGACCTTTTGCTTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTC 1048

Qy   1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
          ||   |    |||| |||| || |||||||   ||||||| |||||     | |||||||
Db   1049 CTCCTTACAATCACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATG 1108

Qy   1205 aaatcatccatctgatg 1221
          ||||  |     |||||
Db   1109 AAATTGTTTCCATGATG 1125


RESULT   4
US-09-073-674-4
; Sequence 4, Application US/09073674
; Patent No. 5998189
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Warner-Lambert Company
;      STREET:  2800 Plymouth Road
;      CITY:  Ann Arbor
;      STATE:  Michigan
;      COUNTRY:  U.S.A.
;      ZIP:  48105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/073,674
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crissey, Todd M.
;      REGISTRATION NUMBER:  37,807
;      REFERENCE/DOCKET NUMBER:  5072-D1-66-TMC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  734 622-7530
;      TELEFAX:  734 622-1553
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1140 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1140
US-09-073-674-4


  Query Match             35.0%;  Score 444.2;  DB 2;  Length 1140;
  Best Local Similarity   62.8%;  Pred. No. 5.6e-139;
  Matches  689;  Conservative    0;  Mismatches  408;  Indels    0;  Gaps    0;

Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
          |||| ||||    |||||||||| ||| | || |||     |    || || ||   |||||
Db     29 ACCCTGAAGTGACCATGAATATAAGTCAGATGATCACCTACTGGGGATACCCAGCTGAGG 88

Qy    185 aatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctcgaggcc 244
          ||||||||||  |  || ||||| || |||||||||     |  |||| ||||||   ||
Db     89 AATATGAAGTTGTGACCGAAGACGGTTATATCCTTGGGATCGACAGAATTCCTTATGGGA 148

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
              |  |  |||| | |||    || ||||| | ||  |||| || ||  |  | |
Db    149 GGAAAAATTCAGAGAATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCG 208

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
            |   || |  |||||||| ||||||||||||||||| ||||||| |||||| |||| |
Db    209 CATCAGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCCTGGCCG 268

Qy    365 atgccggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
          | || || |  |||||||| ||||||||||||||| ||| ||||| |  | |   |
Db    269 ACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACCTGGGCCAGGAGGAATC 328

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
            |     |||    |||   |  || ||||||||||||| | ||| ||||||||||  |
Db    329 TGTACTACTCGCCCGACTCCGTCGAATTCTGGGCTTTCAGCTTTGACGAGATGGCTAAAT 388

Qy    485 ttgaccttcctgcagtgataaactttatttttgcagaaaacgggccaggaaaagatctatt 544
           |||  |||| ||    ||  |||| || ||| ||||||||| ||||| ||| |  | |
Db    389 ATGACCTTCCCGCCACCATTGACTTCATCTTGAAGAAAACGGGACAGGACAAGCTACACT 448

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
          | || ||| |||| |||||||||||||| || || || || ||||||||||  ||  |||
Db    449 ACGTTGGCCATTCCCAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGC 508

Qy    605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
          ||||  |    |||||||  |  | ||| ||||| ||| | ||||| || ||| |  |
Db    509 TGGCCAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGTACACCG 568

Qy    665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
          ||| ||      | |||| |  || |  | ||     |  |  |||||    | ||||| |
Db    569 AAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTCAAGCTTATATTTGGAA 628

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
          | |||  |||       ||   |   |||     |   || |  |   |  | ||   ||
Db    629 ACAAAATATTCTACCCACACCACTTCTTTGATCAATTTCTCGCCACCGAGGTATGCTCCC 688

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
            | ||   | ||||   | || || ||   | ||||  |  |  |||||||  |||| |
Db    689 GCGAGACGGTGGATCTCCTCTGCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTA 748

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
            ||  |||||||||| ||      ||| |||    | || | ||  || ||||||| |
Db    749 TGAACTTGAACATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGG 808

Qy    905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
```

```
Db    809 TTCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAGCTTTTGACT 868

Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
Db    869 GGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGCATGCCTCCCTACTACAACC 928

Qy   1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
Db    929 TGACAGACATGCATGTGCCAATCGCAGTGTGGAACGGTGGCAACGACTTGCTGGCCGACC 988

Qy   1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
Db    989 CTCACGATGTTGACCTTTTGCTTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTC 1048

Qy   1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
Db   1049 CTCCTTACAATCACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATG 1108

Qy   1205 aaatcatccatctgatg 1221
Db   1109 AAATTGTTTCCATGATG 1125


RESULT   5
US-08-227-108-6
; Sequence 6, Application US/08227108
; Patent No. 5807726
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/227,108
;      FILING DATE:  03-APR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fanucci, Allan A.
;      REGISTRATION NUMBER:  30,256
;      REFERENCE/DOCKET NUMBER:  7620-033
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 790-9090
;      TELEFAX:  212 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1146 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
US-08-227-108-6


  Query Match           35.0%;  Score 444.2;  DB 1;  Length 1146;
  Best Local Similarity  62.8%;  Pred. No. 5.6e-139;
  Matches  689;  Conservative    0;  Mismatches  408;  Indels    0;  Gaps    0;

Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
Db     26 ACCCTGAAGTGACCATGAATATAAGTCAGATGATCACCTACTGGGGATACCCAGCTGAGG 85

Qy    185 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 244
Db     86 AATATGAAGTTGTGACCGAAGACGGTTATATCCTTGGGATCGACAGAATTCCTTATGGGA 145

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
Db    146 GGAAAAATTCAGAGAATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCG 205

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
Db    206 CATCAGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCCTGGCCG 265

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
Db    266 ACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACCTGGGCCAGGAGGAATC 325

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
Db    326 TGTACTACTCGCCCGACTCCGTCGAATTCTGGGCTTTCAGCTTTGACGAGATGGCTAAAT 385

Qy    485 ttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 544
Db    386 ATGACCTTCCCGCCACCATTGACTTCATCTTGAAGAAAACGGGACAGGACAAGCTACACT 445

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
Db    446 ACGTTGGCCATTCCCAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGC 505

Qy    605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
Db    506 TGGCGAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGTACACCG 565

Qy    665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
Db    566 AAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTCAAGCTTATATTTGGAA 625

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccttttgtggcc 784
Db    626 ACAAAATATTCTACCCACACCACTTCTTTGATCAATTTCTCGCCACCGAGGTATGCTCCC 685

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
Db    686 GCGAGACGGTGGATCTCCTCTGCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTA 745

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
Db    746 TGAACTTGAACATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGG 805

Qy    905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
Db    806 TTCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAGCTTTTGACT 865

Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
Db    866 GGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGCATGCCTCCCTACTACAACC 925

Qy   1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
Db    926 TGACAGACATGCATGTGCCAATCGCAGTGTGGAACGGTGGCAACGACTTGCTGGCCGACC 985

Qy   1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
Db    986 CTCACGATGTTGACCTTTTGCTTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTC 1045

Qy   1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
Db   1046 CTCCTTACAATCACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATG 1105

Qy   1205 aaatcatccatctgatg 1221
```

```
Db   1106 AAATTGTTTCCATGATG 1122


RESULT   6
US-09-073-674-6
; Sequence 6, Application US/09073674
; Patent No. 5998189
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Warner-Lambert Company
;      STREET:  2800 Plymouth Road
;      CITY:  Ann Arbor
;      STATE:  Michigan
;      COUNTRY:  U.S.A.
;      ZIP:  48105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/073,674
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crissey, Todd M.
;      REGISTRATION NUMBER:  37,807
;      REFERENCE/DOCKET NUMBER:  5072-D1-66-TMC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  734 622-7530
;      TELEFAX:  734 622-1553
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1146 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
US-09-073-674-6


  Query Match           35.0%;  Score 444.2;  DB 2;  Length 1146;
  Best Local Similarity  62.8%;  Pred. No. 5.6e-139;
  Matches  689;  Conservative    0;  Mismatches  408;  Indels    0;  Gaps    0;

Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
          |||| ||||    |||||||| ||| | || |||     |    || || ||   |||||
Db     26 ACCCTGAAGTGACCATGAATATAAGTCAGATGATCACCTACTGGGGATACCCAGCTGAGG 85

Qy    185 aatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctcgaggcc 244
          |||||||||| |  || ||||| || |||||||||     |  |||| ||||||   ||
Db     86 AATATGAAGTTGTGACCGAAGACGGTTATATCCTTGGGATCGACAGAATTCCTTATGGGA 145

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
              |  |  |||| | |||    || ||||| |  ||  |||| || ||   |  | |
Db    146 GGAAAAATTCAGAGAATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCG 205

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
          |   || |  |||||||| |||||||||||||||||| ||||||| |||||| |||||| |
Db    206 CATCAGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCCTGGCCG 265

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
          | || || |  |||||||| |||||||||||||||||| ||| ||||| |  | |   |
Db    266 ACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACCTGGGCCAGGAGGAATC 325

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
            |     |||    |||    |  || |||||||||||||| | ||| ||||||||||  |

Db    326 TGTACTACTCGCCCGACTCCGTCGAATTCTGGGCTTTCAGCTTTGACGAGATGGCTAAAT 385

Qy    485 ttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaaagatctatt 544
          |||||||||| ||    ||  |||| || ||| |||||||||| ||||| ||| |  | |
Db    386 ATGACCTTCCCGCCACCATTGACTTCATCTTGAAGAAAACGGGACAGGACAAGCTACACT 445

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcatttttccaccatgccagagc 604
          | || ||| |||| ||||||||||||||| || || || || ||||||||||  ||  |||
Db    446 ACGTTGGCCATTCCCAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGC 505

Qy    605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
          ||||  |    |||||||  |  | ||| |||| ||| | ||||| || ||| | |
Db    506 TGGCGAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGTACACCG 565

Qy    665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
          ||| ||    | |||| |  || |  |||     |  | |||||    | |||||
Db    566 AAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTCAAGCTTATATTTGGAA 625

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
          | |||  |||        ||   |  |||      |   || | |  |  | | ||  ||
Db    626 ACAAAATATTCTACCCACACCACTTCTTTGATCAATTTCTCGCCACCGAGGTATGCTCCC 685

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
          | ||   | ||||  | || || ||   | ||||  |  |  |||||||  |||| |
Db    686 GCGAGACGGTGGATCTCCTCTGCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTA 745

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
            ||  |||||||||| ||      ||| |||     | || | ||  || |||||||| |
Db    746 TGAACTTGAACATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGG 805

Qy    905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
          | || ||  | || ||||||  |||||| || || |||||  |  |||  || |||||||
Db    806 TTCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAGCTTTTGACT 865

Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
          ||||  ||         | ||  ||     | |  |||||  |  |||       ||||
Db    866 GGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGCATGCCTCCCTACTACAACC 925

Qy   1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
          | | ||| |||   || || | ||| ||||||  || ||  | |||| |||  |  | |
Db    926 TGACAGACATGCATGTGCCAATCGCAGTGTGGAACGGTGGCAACGACTTGCTGGCCGACC 985

Qy   1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
          |  | || ||  |  |  |||| ||  ||  |  |||| |||||  ||||| | ||| ||||
Db    986 CTCAGGATGTTGACCTTTTGCTTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTC 1045

Qy   1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
          ||  |    |||| |||| || |||||||    |||||| |||||     | |||||||
Db    1046 CTCCTTACAATCACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATG 1105

Qy   1205 aaatcatccatctgatg 1221
          ||||  |    |||||
Db    1106 AAATTGTTTCCATGATG 1122


RESULT   7
US-08-227-108-1
; Sequence 1, Application US/08227108
; Patent No. 5807726
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
```

```
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/227,108
;       FILING DATE:  03-APR-1994
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Fanucci, Allan A.
;       REGISTRATION NUMBER:  30,256
;       REFERENCE/DOCKET NUMBER:  7620-033
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  212 790-9090
;       TELEFAX:  212 869-8864/9741
;       TELEX:  66141 PENNIE
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1528 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA to mRNA
US-08-227-108-1


  Query Match           35.0%;  Score 444.2;  DB 1;  Length 1528;
  Best Local Similarity   62.8%;  Pred. No. 6.8e-139;
  Matches  689;  Conservative    0;  Mismatches  408;  Indels    0;  Gaps    0;

Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
          |||| ||||    |||||||| ||| | || |||     |    || || ||   |||||
Db     26 ACCCTGAAGTGACCATGAATATAAGTCAGATGATCACCTACTGGGGATACCCAGCTGAGG 85

Qy    185 aatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctcgaggcc 244
          ||||||||||| |  || ||||| || |||||||||     |  |||| ||||||   ||
Db     86 AATATGAAGTTGTGACCGAAGACGGTTATATCCTTGGGATCGACAGAATTCCTTATGGGA 145

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
             |  | |||| | |||    || ||||| |  ||  |||| || ||  |  | |
Db    146 GGAAAAATTCAGAGAATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCG 205

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
          |  || |  |||||||| |||||||||||||||| ||||||| |||||  |  | |   |
Db    206 CATCAGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCCTGGCCG 265

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
          | || || |  |||||||| |||||||||||||||| ||| ||||| |  | |   |
Db    266 ACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACCTGGGCCAGGAGGAATC 325

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
          |     |||    |||   |  || ||||||||||||| | ||| ||||||||||  |
Db    326 TGTACTACTCGCCCGACTCCGTCGAATTCTGGGCTTTCAGCTTTGACGAGATGGCTAAAT 385

Qy    485 ttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 544
          |||||||| ||    ||  |||| || ||| |||||||||| ||||| ||| |  | |
Db    386 ATGACCTTCCCGCCACCATTGACTTCATCTTGAAGAAAACGGGACAGGACAAGCTACACT 445

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
          | || ||| |||| |||||||||||||| || || || || ||||||||||  ||  |||
Db    446 ACGTTGGCCATTCCCAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGC 505

Qy    605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
          |||| | |    |||||||  |  | ||| ||||| ||| | ||||| || ||| |  |
Db    506 TGGCGAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGTACACCG 565

Qy    665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
          ||| ||     | |||| |  || |  | ||     |  |  |||||    | |||||  |
Db    566 AAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTCAAGCTTATATTTGGAA 625

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
          | |||  |||       ||   |   |||      |   || |  |   |  | ||   ||
Db    626 ACAAAATATTCTACCCACACCACTTCTTTGATCAATTTCTCGCCACCGAGGTATGCTCCC 685

Qy    785 agctgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
          |  ||   | ||||   | || || ||   | ||||  |  |  |||||||  |||| |
Db    686 GCGAGACGGTGGATCTCCTCTGCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTA 745

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
            ||  |||||||||| ||      ||| |||    | || | ||  || |||||||| |
Db    746 TGAACTTGAACATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGG 805

Qy    905 tgcaaaatattctacactggagccaggcagtgaatttctggtgaactccgggcatttgact 964
          | || ||  | || ||||||  |||||| || || |||||  |  |||  || |||||||
Db    806 TTCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAGCTTTTGACT 865

Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
          |||| ||        | ||  ||    | |  |||||   |  |||      ||||
Db    866 GGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGCATGCCTCCCTACTACAACC 925

Qy   1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
          | | ||| |||   || || |  ||| ||||||  || ||  | |||| |||  |  | |
Db    926 TGACAGACATGCATGTGCCAATCGCAGTGTGGAACGGTGGCAACGACTTGCTGGCCGACC 985

Qy   1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
          |  | || ||  |  |  |||| ||  || |  |||| |||| ||||| | ||| ||||
Db    986 CTCACGATGTTGACCTTTTGCTTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTC 1045

Qy   1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
          ||  |    |||| |||| || |||||||   |||||| |||||     | |||||||
Db   1046 CTCCTTACAATCACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATG 1105

Qy   1205 aaatcatccatctgatg 1221
          ||||  |     |||||
Db   1106 AAATTGTTTCCATGATG 1122


RESULT   8
US-09-073-674-1
; Sequence 1, Application US/09073674
; Patent No. 5998189
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Warner-Lambert Company
;      STREET:  2800 Plymouth Road
;      CITY:  Ann Arbor
;      STATE:  Michigan
;      COUNTRY:  U.S.A.
;      ZIP:  48105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/073,674
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crissey, Todd M.
;      REGISTRATION NUMBER:  37,807
;      REFERENCE/DOCKET NUMBER:  5072-D1-66-TMC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  734 622-7530
;      TELEFAX:  734 622-1553
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1528 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
```

```
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
US-09-073-674-1


  Query Match             35.0%;  Score 444.2;  DB 2;  Length 1528;
  Best Local Similarity   62.8%;  Pred. No. 6.8e-139;
  Matches  689;  Conservative    0;  Mismatches  408;  Indels    0;  Gaps    0;

Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
          |||| ||||    |||||||| ||| | || |||    |    || || ||   |||||
Db     26 ACCCTGAAGTGACCATGAATATAAGTCAGATGATCACCTACTGGGGATACCCAGCTGAGG 85

Qy    185 aatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctcgaggcc 244
          |||||||||| |  || ||||| || |||||||||     |  |||| ||||||   ||
Db     86 AATATGAAGTTGTGACCGAAGACGGTTATATCCTTGGGATCGACAGAATTCCTTATGGGA 145

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
            |  | |||| | |||    || ||||| |  ||  |||| || ||  |  | |
Db    146 GGAAAAATTCAGAGAATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCG 205

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
          |   || |  |||||||| |||||||||||||||||| ||||||| |||||| ||||| |
Db    206 CATCAGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCCTGGCCG 265

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
          | || || |  |||||||| |||||||||||||||| ||| ||||| | | |   |
Db    266 ACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACCTGGGCCAGGAGGAATC 325

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
            |     |||    |||    |  || ||||||||||||| | ||| ||||||||||  |
Db    326 TGTACTACTCGCCCGACTCCGTCGAATTCTGGGCTTTCAGCTTTGACGAGATGGCTAAAT 385

Qy    485 ttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 544
           ||||||||| ||    ||  |||| || ||| |||||||||| ||||| ||| |   | |
Db    386 ATGACCTTCCCGCCACCATTGACTTCATCTTGAAGAAAACGGGACAGGACAAGCTACACT 445

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
          | || ||| |||| |||||||||||||| || || || || |||||||||  ||  |||
Db    446 ACGTTGGCCATTCCCAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGC 505

Qy    605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
          ||||  |    |||||||  |  | ||| |||| ||| | |||||  || ||| |  |
Db    506 TGGCGAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGTACACCG 565

Qy    665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
          ||| ||     | |||| |  || |  | ||     |  |  |||||    | ||||| |
Db    566 AAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTCAAGCTTATATTTGGAA 625

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccttttgtggcc 784
          | |||  |||        ||   |  |||      |  || |  |   |  | ||  ||
Db    626 ACAAAATATTCTACCCACACCACTTCTTTGATCAATTTCTCGCCACCGAGGTATGCTCCC 685

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
            | ||   | ||||   | || || ||   | ||||  |  |  |||||||  |||| |
Db    686 GCGAGACGGTGGATCTCCTCTGCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTA 745

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
            ||  |||||||||| ||      ||| |||    | || | ||  || |||||||| |
Db    746 TGAACTTGAACATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGG 805

Qy    905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
          | || ||  | || ||||||  |||||| || || |||||  |  |||  || |||||||
Db    806 TTCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAGCTTTTGACT 865

Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
          ||||  ||         | ||  ||     | |   |||||    |  |||        ||||
Db    866 GGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGCATGCCTCCCTACTACAACC 925

Qy   1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
          | | ||| |||   || || |  ||| ||||||  || ||  | |||| |||  |  | |
Db    926 TGACAGACATGCATGTGCCAATCGCAGTGTGGAACGGTGGCAACGACTTGCTGGCCGACC 985

Qy   1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
          |  | || ||  |  |  |||| ||  || |  |||| ||||| ||||| | ||| ||||
Db    986 CTCACGATGTTGACCTTTTGCTTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTC 1045

Qy   1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
          ||   |   |||| |||| || |||||||   |||||| |||||     | |||||||
Db    1046 CTCCTTACAATCACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATG 1105

Qy   1205 aaatcatccatctgatg 1221
          |||'  |     |||||
Db    1106 AAATTGTTTCCATGATG 1122


RESULT   9
US-08-751-782-2
; Sequence 2, Application US/08751782
; Patent No. 5821352
;  GENERAL INFORMATION:
;    APPLICANT:  Heintz, Nathaniel
;    APPLICANT:  Gubbay, Johnathan
;    APPLICANT:  Skinner, Michael
;    TITLE OF INVENTION:  A cDNA Library Prepared during
;    TITLE OF INVENTION:  Regression of Rat Prostate and Methods of Use Thereof
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  David A. Jackson, Esq.
;      STREET:  411 Hackensack Ave, Continental Plaza, 4th
;      STREET:  Floor
;      CITY:  Hackensack
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/751,782
;      FILING DATE:  18-NOV-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jackson Esq., David A.
;      REGISTRATION NUMBER:  26,742
;      REFERENCE/DOCKET NUMBER:  600-1-190
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-487-5800
;      TELEFAX:  201-343-1684
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  178 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    IMMEDIATE SOURCE:
;      CLONE:  10.2
US-08-751-782-2


  Query Match             3.8%;  Score 48.2;  DB 1;  Length 178;
  Best Local Similarity   55.8%;  Pred. No. 1.4e-06;
  Matches   92;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy    416 gaaaacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgaga 475
          | |||||   |   ||   |  |||     | || |||| ||| |||| | ||||
Db      8 GGAAACATGTGCGCCTAGACCCAGATTCTAAAGAATTTTGGGATTTTAGTTTTAATGAAC 67

Qy    476 tggctaggtttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaa 535
                |  ||||| || ||  | ||  | || ||| || |  ||||  | ||   |
Db     68 AAATAGAATACGACCTCCCAGCCATCATTTATTTCATTCTGAATGAAACAAGACAAACAC 127
```

```
Qy    536 agatctattatgtcggctattcacagggcaccaccatgggcttta 580
          | || || ||| | ||| |||| ||||||  |    || |  | |
Db    128 AAATTTACTATATTGGCCATTCCCAGGGCGTCTATCTGCGTATCA 172


RESULT 10
US-08-925-171-2
; Sequence 2, Application US/08925171
; Patent No. 5928871
;  GENERAL INFORMATION:
;    APPLICANT:  Heintz, Nathaniel
;    APPLICANT:  Gubbay, Johnathan
;    APPLICANT:  Skinner, Michael
;    TITLE OF INVENTION:  A cDNA Library Prepared during
;    TITLE OF INVENTION:  Regression of Rat Prostate and Methods of Use Thereof
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  David A. Jackson, Esq.
;      STREET:  411 Hackensack Ave, Continental Plaza, 4th
;      STREET:  Floor
;      CITY:  Hackensack
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/925,171
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jackson Esq., David A.
;      REGISTRATION NUMBER:  26,742
;      REFERENCE/DOCKET NUMBER:  600-1-190 CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-487-5800
;      TELEFAX:  201-343-1684
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  178 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    IMMEDIATE SOURCE:
;      CLONE:  10.2
US-08-925-171-2


  Query Match             3.8%;  Score 48.2;  DB 2;  Length 178;
  Best Local Similarity   55.8%;  Pred. No. 1.4e-06;
  Matches   92;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy    416 gaaaacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgaga 475
          | |||||   |    ||   |  |||     | || || |||| ||| |||| | ||||
Db      8 GGAAACATGTGCGCCTAGACCCAGATTCTAAAGAATTTTGGGATTTTAGTTTTAATGAAC 67

Qy    476 tggctaggtttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaa 535
              |  ||||| || ||  | ||  | || ||| || |  ||||  | ||   |
Db     68 AAATAGAATACGACCTCCCAGCCATCATTTATTTCATTCTGAATGAAACAAGACAAACAC 127

Qy    536 agatctattatgtcggctattcacagggcaccaccatgggcttta 580
          | || || ||| | ||| |||| ||||||  |    || |  | |
Db    128 AAATTTACTATATTGGCCATTCCCAGGGCGTCTATCTGCGTATCA 172


RESULT 11
US-09-356-952-12
; Sequence 12, Application US/09356952
; Patent No. 6117663
; GENERAL INFORMATION:
;  APPLICANT: Boriack-Sjodin, Ann
;  APPLICANT: Margarit, S. M.
;  APPLICANT: Bor-Sogi, Dafna
;  APPLICANT: Cole, Philip
;  APPLICANT: Kuriyan, John
;  TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE
;  TITLE OF INVENTION: THEREOF
;  FILE REFERENCE: 600-1-228N
;  CURRENT APPLICATION NUMBER: US/09/356,952
;  CURRENT FILING DATE: 1999-07-19
;  EARLIER APPLICATION NUMBER: 60/093,631
;  EARLIER FILING DATE: 1998-07-21
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
;   LENGTH: 43676
;   TYPE: DNA
;   ORGANISM: Saccharomyces cerevisiae
US-09-356-952-12


  Query Match             3.7%;  Score 46.4;  DB 3;  Length 43676;
  Best Local Similarity   46.0%;  Pred. No. 0.0002;
  Matches  232;  Conservative    0;  Mismatches  266;  Indels    6;  Gaps    2;

Qy     94 aattcagtacatatgccaactaaagctgtggacccagaagcattcatgaatattagtgaa 153
          ||| || |||| |  | ||    | | |  |||     |  | |   | || ||| | |
Db  14682 aatccattacagaatctaatcttatcggaagacttgacattagttgcggatcttaattat 14741

Qy    154 atcatccaacatcaaggctatccctgtgaggaatatgaagtcgcaactgaagatgggt-- 211
              |  | ||  |      ||     |||||||  |  | | |  || ||||||||||
Db  14742 tattttaaccagtacaatattcaaatagaggaatttagattggagaccgaagatgggttt 14801

Qy    212 -atatcctttctgttaacaggattcctcgaggcctagtgcaacctaagaagacaggttcc 270
           |||   ||      ||  ||| ||   |     |  |  |  |    | |
Db  14802 gttatagatttgtggcacttgataccaaaatatagaacgacagattctgacaagaagaag 14861

Qy    271 aggcctgtggtgttactgcagcatggcctagttggaggtgctagcaactggatttccaac 330
          |||||     | ||  |||  ||||||||  |   | |   | |     |     |||||
Db  14862 aggccacccattttgatgctacatggccttttgcaaagcagtggttcgttcgcatccaa- 14920

Qy    331 ctgcccaacaatagcctgggcttcattctggcagatgctggttttgacgtgtggatgggg 390
              |  ||    ||||  |   |||||    |  | ||||  ||| | ||| | |||
Db  14921 --tggtagaaaatctctggcatattttctgtatcaatccggttacgacatatggttaggg 14978

Qy    391 aacagcaggggaaacgcctggtctcgaaaacacaagacactctccatagaccaagatgag 450
          || | |||  |        || |   |    || |  |                | || |
Db  14979 aataacagatgcgggtttaggccggaatggaacgaagcgaaagtaccgacactagcttcc 15038

Qy    451 ttctgggctttcagttatgatgagatggctaggtttgaccttcctgcagtgataaacttt 510
            ||||  |       |  |||||||  || ||  || ||  |    ||||  |
Db  15039 aggtgggactgggaccttcgcgagatggttaagtacgatctgacccttttgattgatacc 15098

Qy    511 attttgcagaaaacgggccaggaaaagatctattatgtcggctattcacagggcaccacc 570
            | ||   || |||      ||||||  |   ||   ||    |||| |||||||| ||
Db  15099 gtgttagctaagacgcagtttgaaaagcttactttgatctcgcattctcagggcactaca 15158

Qy    571 atgggctttattgcattttccacc 594
            ||| ||||| |  ||   || |
Db  15159 caggggtttatgggcttggtcaac 15182


RESULT 12
US-08-852-806-1
; Sequence 1, Application US/08852806
; Patent No. 5874245
;  GENERAL INFORMATION:
;    APPLICANT:  Shoji Fukusumi
;    APPLICANT:  Shuji Hinuma
;    TITLE OF INVENTION:  NOVEL HUMAN G-PROTEIN COUPLED RECEPTOR
;    NUMBER OF SEQUENCES:  14
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  U.S.A.
;      ZIP:  19406-0939
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/852,806
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/017,915
;      FILING DATE:  16 MAY 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Han, William T
;      REGISTRATION NUMBER:  34,344
;      REFERENCE/DOCKET NUMBER:  TAK50002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-5219
;      TELEFAX:  610-270-5090
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3271 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
US-08-852-806-1


  Query Match             2.9%;  Score 36.6;  DB 2;  Length 3271;
  Best Local Similarity   50.9%;  Pred. No. 0.074;
  Matches   87;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy     969 gagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagagtcag 1028
           || ||  ||  | ||||| |  |   ||| ||||| |   || ||           |||
Db     678 GATTGTCACAGATAATCTAGTGATCAGCATTCAGCGAGAGCCCGTCTCAGCTGTGTCCAG 737

Qy    1029 agatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatccaga 1088
           || || ||| |||| |          |    ||     |||||||| |     |  ||||
Db     738 TGACATCACGTTCCCCATGCGGGGCCGCCGGGGCATGAAGGACTGGGTGCGGCACTCAGA 797

Qy    1089 agacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaa 1139
           |||       | |||| |   ||||||  || ||||  || ||    |||
Db     798 GGACCGCCTCTTCCTGCCCAAGGAGGTGCTCAGCCTCTCCTCCCCAGGGAA 848


RESULT  13
US-09-163-669-1
; Sequence 1, Application US/09163669
; Patent No. 6111076
;  GENERAL INFORMATION:
;    APPLICANT:  FUKUSUMI, SHOJI
;    APPLICANT:  HINUMA, SHUJI
;    APPLICANT:  FUJII, RYO
;    TITLE OF INVENTION:  NOVEL HUMAN G-PROTEIN COUPLED
;    TITLE OF INVENTION:  RECEPTOR (HIBCD07)
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ratner & Prestia
;      STREET:  P.O. Box 980
;      CITY:  Valley Forge
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19482
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/163,669
;      FILING DATE:  30-SEP-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/852,806
;      FILING DATE:  07-MAY-1997
;      APPLICATION NUMBER:  60/017,915
;      FILING DATE:  16-MAY-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Prestia, Paul F
;      REGISTRATION NUMBER:  23,031
;      REFERENCE/DOCKET NUMBER:  TAK-50002-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-407-0700
;      TELEFAX:  610-407-0700
;      TELEX:  846169
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3271 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-163-669-1


  Query Match             2.9%;  Score 36.6;  DB 3;  Length 3271;
  Best Local Similarity   50.9%;  Pred. No. 0.074;
  Matches   87;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy     969 gagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagagtcag 1028
           || ||  ||  | ||||| |  |   ||| ||||| |   || ||           |||
Db     678 GATTGTCACAGATAATCTAGTGATCAGCATTCAGCGAGAGCCCGTCTCAGCTGTGTCCAG 737

Qy    1029 agatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatccaga 1088
           || || ||| |||| |          |    ||     |||||||| |     |  ||||
Db     738 TGACATCACGTTCCCCATGCGGGGCCGCCGGGGCATGAAGGACTGGGTGCGGCACTCAGA 797

Qy    1089 agacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaa 1139
           |||       | |||| |   ||||||  || ||||  || ||    |||
Db     798 GGACCGCCTCTTCCTGCCCAAGGAGGTGCTCAGCCTCTCCTCCCCAGGGAA 848


RESULT  14
US-08-276-852-154
; Sequence 154, Application US/08276852
; Patent No. 5652138
;  GENERAL INFORMATION:
;    APPLICANT:  Burton, Dennis R
;    APPLICANT:  Barbas, Carlos F
;    APPLICANT:  Lerner, Richard A
;    TITLE OF INVENTION:  HUMAN NEUTRALIZING MONOCLONAL ANTIBODIES
;    TITLE OF INVENTION:  TO HUMAN IMMUNODEFICIENCY VIRUS
;    NUMBER OF SEQUENCES:  170
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  The Scripps Research Institute, Office of
;      ADDRESSEE:  Patent Counsel
;      STREET:  10666 No. 5652138th Torrey Pines Road, Suite 220,
;      STREET:  Mail Drop TPC8
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
```

```
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/276,852
;      FILING DATE:  18-JUL-1994
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/178,302
;      FILING DATE:  30-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/954,148
;      FILING DATE:  30-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fitting, Thomas
;      REGISTRATION NUMBER:  34,163
;      REFERENCE/DOCKET NUMBER:  SCR1452P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-554-2937
;      TELEFAX:  619-554-6312
;  INFORMATION FOR SEQ ID NO:  154:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3282 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  15..452
US-08-276-852-154


  Query Match             2.9%;  Score 36.6;  DB 1;  Length 3282;
  Best Local Similarity   51.5%;  Pred. No. 0.074;
  Matches   84;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy    435 catagaccaagatgagttctgggctttcagttatgatgagatggctaggtttgaccttcc 494
          || || | |    |||||  ||    |||| | ||  || | ||||   | | |   | |
Db    302 CACAGCCTACATGGAGTTGAGGAGCCTCAGGTCTGCAGACACGGCTGTTTATTATTGTGC 361

Qy    495 tgcagtgataaactttattttgcagaaaacgggccaggaaaagatctattatgtcggcta 554
            ||||      | || || | |  | |   |||||| ||    |||   | || ||
Db    362 GAGAGTGGGGCCATATAGTTGGGATGATTCTCCCCAGGACAATTATTATATGGACGTCTG 421

Qy    555 ttcacagggcaccaccatgggctttattgcattttccaccatg 597
              | || |||||  |   | | |   ||  |||||||| |
Db    422 GGGCAAAGGAACCACGGTCATCGTGAGCTCAGCTTCCACCAAG 464


RESULT  15
US-08-276-852-169/c
; Sequence 169, Application US/08276852
; Patent No. 5652138
;  GENERAL INFORMATION:
;    APPLICANT:  Burton, Dennis R
;    APPLICANT:  Barbas, Carlos F
;    APPLICANT:  Lerner, Richard A
;    TITLE OF INVENTION:  HUMAN NEUTRALIZING MONOCLONAL ANTIBODIES
;    TITLE OF INVENTION:  TO HUMAN IMMUNODEFICIENCY VIRUS
;    NUMBER OF SEQUENCES:  170
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  The Scripps Research Institute, Office of
;      ADDRESSEE:  Patent Counsel
;      STREET:  10666 No. 5652138th Torrey Pines Road, Suite 220,
;      STREET:  Mail Drop TPC8
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/276,852
;      FILING DATE:  18-JUL-1994
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/178,302
;      FILING DATE:  30-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/954,148
;      FILING DATE:  30-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fitting, Thomas
;      REGISTRATION NUMBER:  34,163
;      REFERENCE/DOCKET NUMBER:  SCR1452P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-554-2937
;      TELEFAX:  619-554-6312
;  INFORMATION FOR SEQ ID NO:  169:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3282 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-276-852-169


  Query Match             2.9%;  Score 36.6;  DB 1;  Length 3282;
  Best Local Similarity   51.5%;  Pred. No. 0.074;
  Matches   84;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy    435 catagaccaagatgagttctgggctttcagttatgatgagatggctaggtttgaccttcc 494
          || || | |    |||||  ||    |||| | ||  || | ||||   | | |   | |
Db   2981 CACAGCCTACATGGAGTTGAGGAGCCTCAGGTCTGCAGACACGGCTGTTTATTATTGTGC 2922

Qy    495 tgcagtgataaactttattttgcagaaaacgggccaggaaaagatctattatgtcggcta 554
            ||||      | || || | |  | |   |||||| ||    |||   | || ||
Db   2921 GAGAGTGGGGCCATATAGTTGGGATGATTCTCCCCAGGACAATTATTATATGGACGTCTG 2862

Qy    555 ttcacagggcaccaccatgggctttattgcattttccaccatg 597
              | || |||||  |   | | |   ||  |||||||| |
Db   2861 GGGCAAAGGAACCACGGTCATCGTGAGCTCAGCTTCCACCAAG 2819


Search completed: August 24, 2001, 22:39:52
Job time: 7308 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: August 24, 2001, 22:44:11 ; Search time 4395.77 Seconds
(without alignments)
4465.332 Million cell updates/sec

Title: US-09-333-159-46
Perfect score: 1269
Sequence: 1 atgttggaaaccttgtcaag.....gacggtgtgaggccgtattg 1269

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	527.8	41.6	2481	93	HSLYACLY	Z31690 H.sapiens (
	2	527.8	41.6	2481	97	HSU08464	U08464 Human lysos
	3	527.8	41.6	2626	93	HSLAL	X76488 H.sapiens m
	4	526.2	41.5	2493	97	HUMLIPCHL	M74775 Human lysos
	5	489.2	38.6	1378	9	A26689	A26689 Precursor o
C	6	489.2	38.6	1378	9	A26690	A26690 Precursor o
	7	470	37.0	1365	93	HSGLR	X05997 Human mRNA
	8	470	37.0	1367	9	A01046	A01046 H.sapiens m

	9	470	37.0	1367	9	A12714 A12714 Pregastric
	10	467.6	36.8	1336	9	A01157 A01157 R.norvegicu
	11	467.6	36.8	1355	95	RNLIP X02309 Rat mRNA fo
	12	461	36.3	2358	94	MMLYACLY Z31689 M.musculus
	13	459.8	36.2	3144	95	S81497 S81497 lysosomal a
	14	458	36.1	1417	7	BOVMPE L26319 Bos taurus
	15	444.2	35.0	1137	9	AR039022 AR039022 Sequence
	16	444.2	35.0	1137	9	AR092633 AR092633 Sequence
	17	444.2	35.0	1140	9	AR039023 AR039023 Sequence
	18	444.2	35.0	1140	9	AR092634 AR092634 Sequence
	19	444.2	35.0	1146	9	AR039024 AR039024 Sequence
	20	444.2	35.0	1146	9	AR092635 AR092635 Sequence
	21	444.2	35.0	1198	9	A57760 A57760 Sequence 5
	22	444.2	35.0	1528	9	A57756 A57756 Sequence 1
	23	444.2	35.0	1528	9	AR039021 AR039021 Sequence
	24	444.2	35.0	1528	9	AR092632 AR092632 Sequence
	25	444.2	35.0	1651	7	CFLIPASE Y13899 Canis famil
	26	441	34.8	1137	9	A39301 A39301 Sequence 2
	27	441	34.8	1140	9	A39303 A39303 Sequence 4
	28	441	34.8	1146	9	A39305 A39305 Sequence 6
	29	441	34.8	1528	9	A39300 A39300 Sequence 1
	30	388.8	30.6	1048	9	A57758 A57758 Sequence 3
	31	286	22.5	699	10	AX068249 AX068249 Sequence
	32	286	22.5	1308	10	AX068259 AX068259 Sequence
c	33	269	21.2	76653	63	AC015506 AC015506 Homo sapi
c	34	269	21.2	158195	63	AC015461 AC015461 Homo sapi
	35	269	21.2	181170	79	AL353113 AL353113 Homo sapi
	36	204.4	16.1	148720	65	AC019353 AC019353 Homo sapi
	37	204.4	16.1	172834	62	AC011277 AC011277 Homo sapi
	38	151.2	11.9	76653	63	AC015506 AC015506 Homo sapi
	39	147	11.6	161826	90	AL358532 AL358532 Human DNA
	40	140.6	11.1	401	93	HSLIPCHL04 U04288 Human lysos
	41	140.6	11.1	1851	93	HSLIPA4 X75491 H.sapiens L
c	42	140.6	11.1	180366	81	AL513533 AL513533 Homo sapi
	43	127.6	10.1	40114	6	CELZK6 AF067942 Caenorhab
	44	122.4	9.6	137911	61	AC009769 AC009769 Homo sapi
	45	122.4	9.6	204102	63	AC013826 AC013826 Homo sapi

ALIGNMENTS

```
RESULT   1
HSLYACLY
LOCUS       HSLYACLY     2481 bp    mRNA            PRI       17-FEB-1997
DEFINITION  H.sapiens (HepG2) LAL mRNA for lysosomal acid lipase.
ACCESSION   Z31690
VERSION     Z31690.1  GI:506430
KEYWORDS    LAL; lipase; lysosomal acid lipase.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2481)
  AUTHORS   Du,H. and Gregory,G.A.
  TITLE     Structural Conservation of Putative Functional Motifs between Mouse
            and Human Lysosomal Acid Lipase
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2481)
  AUTHORS   Du,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-APR-1994) Hong Du, Division of Human Genetics,
            Children's Hospital Medical Center, 3333 Bernet Street, Cincinnati,
            OH, 45229-3039, USA
FEATURES             Location/Qualifiers
     source          1. .2481
                     /organism="Homo sapiens"
                     /strain="HepG2"
                     /db_xref="taxon:9606"
                     /clone="clone hLAL"
                     /tissue_type="liver"
                     /cell_type="hepatoma"
                     /clone_lib="human liver HepG2 cDNA library/lambda ZAP"
     sig_peptide     34. .96
                     /gene="LAL"
     gene            34. .1233
                     /gene="LAL"
     CDS             34. .1233
                     /gene="LAL"
                     /citation=[1]
                     /codon_start=1
                     /product="lysosomal acid lipase"
                     /protein_id="CAA83495.1"
                     /db_xref="GI:506431"
                     /db_xref="SPTREMBL:Q16529"
                     /translation="MKMRFLGLVVCLVLWTLHSEGSRGKLTAVDPETNMNVSEIISYW
                     GFPSEEYLVETEDGYILCLNRIPHGRKNHSDKGPKPVVFLQHGLLADSSNWVTNLANS
                     SLGFILADAGFDVWMGNSRGNTWSRKHKTLSVSQDEFWAFSYDEMAKYDLPASINFIL
                     NKTGQEQVYYVGHSQGTTIGFIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGR
                     LPDHLIKDLFGDKEFLPQSAFLKWLGTHVCTHVILKELCGNLCFLLCGFNERNLNMSR
                     VDVYTTHSPAGTSVQNMLHWSQAVKFQKFQAFDWGSSAKNYFHYNQSYPPTYNVKDML
                     VPTAVWSGGHDWLADVYDVNILLTQITNLVFHESIPEWEHLDFIWGLDAPWRLYNKII
                     NLMRKYQ"
     variation       79
                     /gene="LAL"
                     /citation=[1]
                     /replace="c"
     mat_peptide     97. .1230
                     /gene="LAL"
                     /product="lysosomal acid lipase"
     variation       100
                     /gene="LAL"
                     /citation=[1]
                     /replace="g"
     polyA_signal    2459. .2464
     polyA_site      2480
BASE COUNT      692 a    477 c    513 g    799 t
ORIGIN
```

Query Match 41.6%; Score 527.8; DB 93; Length 2481;
Best Local Similarity 65.4%; Pred. No. 4.5e-146;
Matches 791; Conservative 0; Mismatches 412; Indels 6; Gaps 1;

```
Qy      39 cagaatggaaatgtggcttctgattctggtggcgtatatgttccagagaaatgtgaattc 98
           ||||||| ||||| || |  ||    ||||||  | | || | |   | |   || ||||
Db      30 CAGAATGAAAATGCGGTTCTTGGGGTTGGTGGTCTGTTTGGTTCTCTGGACCCTGCATTC 89

Qy      99 ag------tacatatgccaactaaagctgtggacccagaagcattcatgaatattagtga 152
           ||      |  |   | |||||||||| || ||| ||  |||||||  | |||||
Db      90 TGAGGGGTCTAGAGGGAAACTGACAGCTGTGGATCCTGAAACAAACATGAATGTGAGTGA 149

Qy     153 aatcatccaacatcaaggctatccctgtgaggaatatgaagtcgcaactgaagatgggta 212
           ||| |||    |   || |  ||  ||||||||||   ||| |  || |||||||| ||
Db     150 AATTATCTCTTACTGGGGATTCCCTAGTGAGGAATACCTAGTTGAGACAGAAGATGGATA 209

Qy     213 tatcctttctgttaacaggattcctcgaggcctagtgcaacctaagaagacaggttccag 272
           ||| || |  ||||| | |||||||  ||      | | | |   | | |||| |||
Db     210 TATTCTGTGCCTTAACCGAATTCCTCATGGGAGGAAGAACCATTCTGACAAAGGTCCCAA 269

Qy     273 gcctgtggtgttactgcagcatggcctagttggaggtgctagcaactggatttccaacct 332
           || || || || ||||| |||||| |  | | || | |||| |||||| |  | |||||
Db     270 ACCAGTTGTCTTCCTGCAACATGGCTTGCTGGCAGATTCTAGTAACTGGGTCACAAACCT 329

Qy     333 gcccaacaatagcctgggcttcattctggcagatgctggttttgacgtgtggatggggaa 392
           ||||||  |||||||||||||||| || |||||||||||||||||||||||||||||| ||
Db     330 TGCCAACAGCAGCCTGGGCTTCATTCTTGCTGATGCTGGTTTTGACGTGTGGATGGGCAA 389

Qy     393 cagcaggggaaacgcctggtctcgaaaacacaagacactctccatagaccaagatgagtt 452
           ||||  | |||||  |||||||||| ||||| ||||||||||  |    || ||||| ||
Db     390 CAGCAGAGGAAATACCTGGTCTCGGAAACATAAGACACTCTCAGTTTCTCAGGATGAATT 449

Qy     453 ctgggctttcagttatgatgagatggctaggtttgaccttcctgcagtgataaactttat 512
           ||||||||||||||||||||||||||| |  | |||||| || ||    || ||||| ||
Db     450 CTGGGCTTTCAGTTATGATGAGATGGCAAAATATGACCTACCAGCTTCCATTAACTTCAT 509
```

```
Qy    513 tttgcagaaaacgggccaggaaaagatctattatgtcggctattcacagggcaccaccat 572
          | || | ||||| ||||| ||| |  | ||||||||| ||  |||| || |||||||| ||
Db    510 TCTGAATAAAACTGGCCAAGAACAAGTGTATTATGTGGGTCATTCTCAAGGCACCACTAT 569

Qy    573 gggctttattgcattttccaccatgccagagctggctcagaaaatcaaaatgtattttgc 632
           || ||||| |||||||||    || || |||||||||| | |  || |||||||| ||||||
Db    570 AGGTTTTATAGCATTTTCACAGATCCCTGAGCTGGCTAAAAGGATTAAAATGTTTTTTGC 629

Qy    633 tttagcacccatagccactgttaagcatgcaaaaagccccgggaccaaatttttgttgct 692
          | |  ||  | ||  | ||          |  |||||  | |||||||        |
Db    630 CCTGGGTCCTGTGGCTTCCGTCGCCTTCTGTACTAGCCCTATGGCCAAATTAGGACGATT 689

Qy    693 gccagatatgatgatcaagggattgtttggcaaaaaagaatttctgtatcagaccagatt 752
          ||||||    | || ||||  || |||||  | ||||||||||·   ||||     ||
Db    690 ACCAGATCATCTCATTAAGGACTTATTTGGAGACAAAGAATTTCTTCCCCAGAGTGCGTT 749

Qy    753 tctcagacaacttgttatttacctttgtggccaggtgattcttgatcagatttgtagtaa 812
          | | |     || | ||   || ||||    || || || ||  |  || | ||| | ||
Db    750 TTTGAAGTGGCTGGGTACCCACGTTTGCACTCATGTCATACTGAAGGAGCTCTGTGGAAA 809

Qy    813 tatcatgttacttctgggtggattcaacaccaacaatatgaacatgagccgagcaagtgt 872
          | ||   || |||||| ||||||| ||    |  ||| | || |||   |||    |||
Db    810 TCTCTGTTTTCTTCTGTGTGGATTTAATGAGAGAAATTTAAATATGTCTAGAGTGGATGT 869

Qy    873 atatgctgcccacactcttgctggaacatctgtgcaaaatattctacactggagccaggc 932
          |||| |  | ||  ||| |||||||||| |||||||||| ||  ||||||||||||||||
Db    870 ATATACAACACATTCTCCTGCTGGAACTTCTGTGCAAAACATGTTACACTGGAGCCAGGC 929

Qy    933 agtgaattctggtgaactccgggcatttgactgggggagtgagaccaaaaatctggaaaa 992
          || || |       |  | |  || ||||||||||| ||     |||| |||       |
Db    930 TGTTAAATTCCAAAAGTTTCAAGCCTTTGACTGGGGAAGCAGTGCCAAGAATTATTTTCA 989

Qy    993 atgcaatcagccaactcctgtaaggtacagagtcagagatatgacggtccctacagcaat 1052
          | ||| |||     ||||   |  ||||  || |  || |||   || || || ||| |
Db    990 TTACAACCAGAGTTATCCTCCCACATACAATGTGAAGGACATGCTTGTGCCGACTGCAGT 1049

Qy   1053 gtggacaggaggtcaggactggctttcaaatccagaagacgtgaaaatgctgctctctga 1112
          ||||  || ||||| ||||||||| || ||    | ||||| || ||  | ||  || |
Db   1050 CTGGAGCGGGGGTCACGACTGGCTTGCAGATGTCTACGACGTCAATATCTTACTGACTCA 1109

Qy   1113 ggtgaccaacctcatctaccataagaatattcctgaatgggctcacgtggatttcatctg 1172
          | | |||||| |  | | |||| |||  ||||| |||||||  || · | || ||||| ||
Db   1110 GATCACCAACTTGGTGTTCCATGAGAGCATTCCGGAATGGGAGCATCTTGACTTCATTTG 1169

Qy   1173 gggtttggatgctcctcaccgtatgtacaatgaaatcatccatctgatgcagcaggagga 1232
          |||  ||||||| |||    |  | || ||| |||| ||  |||| |||  | |  |  |
Db   1170 GGGCCTGGATGCCCCTTGGAGGCTTTATAATAAAATTATTAATCTAATGAGGAAATATCA 1229

Qy   1233 gaccaacct 1241
          |   || ||
Db   1230 GTGAAAGCT 1238
```

```
RESULT   2
HSU08464
LOCUS       HSU08464     2481 bp    mRNA            PRI       23-JUN-1994
DEFINITION  Human lysosomal acid lipase mRNA, complete cds.
ACCESSION   U08464
VERSION     U08464.1  GI:505052
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2481)
  AUTHORS   Du,H. and Gregory,G.A.
  TITLE     Structural Conservation of Putative Functional Motifs between Mouse
            and Human Lysosomal Acid Lipase
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2481)
  AUTHORS   Du,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-APR-1994) Hong Du, Division of Human Gentics,
            Children's Hospital Medical Center, 3333 Bernet Street, Cincinnati,
            OH 45229-3039, USA
FEATURES             Location/Qualifiers
     source          1. .2481
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="clone hLAL"
                     /clone_lib="HepG2 cDNA library/Lambda ZAP"
                     /cell_line="HepG2"
                     /cell_type="hepatoma"
                     /tissue_type="liver"
     5'UTR           1. .33
     sig_peptide     34. .96
     CDS             34. .1233
                     /codon_start=1
                     /product="lysosomal acid lipase"
                     /protein_id="AAB60328.1"
                     /db_xref="GI:505053"
                     /translation="MKMRFLGLVVCLVLWTLHSEGSRGKLTAVDPETNMNVSEIISYW
                     GFPSEEYLVETEDGYILCLNRIPHGRKNHSDKGPKPVVFLQHGLLADSSNWVTNLANS
                     SLGFILADAGFDVWMGNSRGNTWSRKHKTLSVSQDEFWAFSYDEMAKYDLPASINFIL
                     NKTGQEQVYYVGHSQGTTIGFIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGR
                     LPDHLIKDLFGDKEFLPQSAFLKWLGTHVCTHVILKELCGNLCFLLCGFNERNLNMSR
                     VDVYTTHSPAGTSVQNMLHWSQAVKFQKFQAFDWGSSAKNYFHYNQSYPPTYNVKDML
                     VPTAVWSGGHDWLADVYDVNILLTQITNLVFHESIPEWEHLDFIWGLDAPWRLYNKII
                     NLMRKYQ"
     variation       79
                     /replace="c"
     mat_peptide     97. .1230
                     /product="lysosomal acid lipase"
     variation       100
                     /replace="g"
     3'UTR           1234. .2480
     polyA_signal    2459. .2464
     polyA_site      2480
BASE COUNT      692 a    477 c    513 g    799 t
ORIGIN
```

```
  Query Match             41.6%;  Score 527.8;  DB 97;  Length 2481;
  Best Local Similarity   65.4%;  Pred. No. 4.5e-146;
  Matches  791;  Conservative    0;  Mismatches  412;  Indels    6;  Gaps    1;

Qy      39 cagaatggaaatgtggcttctgattctggtggcgtatatgttccagagaaatgtgaattc 98
           ||||||| ||||| || | |  ||   . ||||||  | | || | |   | |   || ||||
Db      30 CAGAATGAAAATGCGGTTCTTGGGGTTGGTGGTCTGTTTGGTTCTCTGGACCCTGCATTC 89

Qy      99 ag------tacatatgccaactaaagctgtggacccagaagcattcatgaatattagtga 152
           |       |  |   | |||||||| || ||| ||  ||||||| | |||||
Db      90 TGAGGGGTCTAGAGGGAAACTGACAGCTGTGGATCCTGAAACAAACATGAATGTGAGTGA 149

Qy     153 aatcatccaacatcaaggctatccctgtgaggaatatgaagtcgcaactgaagatgggta 212
           ||| |||    |    || |  ||  ||||||||||   ||| |  || |||||||| ||
Db     150 AATTATCTCTTACTGGGGATTCCCTAGTGAGGAATACCTAGTTGAGACAGAAGATGGATA 209

Qy     213 tatcctttctgttaacaggattcctcgaggcctagtgcaacctaagaagacaggttccag 272
           ||| || |   ||||| | ||||||  ||       | | | |    | | |||| |||
Db     210 TATTCTGTGCCTTAACCGAATTCCTCATGGGAGGAAGAACCATTCTGACAAAGGTCCCAA 269

Qy     273 gcctgtggtgttactgcagcatggcctagttggaggtgctagcaactggatttccaacct 332
           ||  | || || ||||| ||||||  |  | | || | |||| ||||||  |  | |||||
Db     270 ACCAGTTGTCTTCCTGCAACATGGCTTGCTGGCAGATTCTAGTAACTGGGTCACAAACCT 329

Qy     333 gcccaacaatagcctgggcttcattctggcagatgctggttttgacgtgtggatggggaa 392
           ||||||  |||||||||||||||||| || |||||||||||||||||||||||||||| ||
Db     330 TGCCAACAGCAGCCTGGGCTTCATTCTTGCTGATGCTGGTTTTGACGTGTGGATGGGCAA 389

Qy     393 cagcaggggaaacgcctggtctcgaaaacacaagacactctccatagaccaagatgagtt 452
           |||||| |||||  |||||||||| ||||| ||||||||||  |    || |||||  ||
Db     390 CAGCAGAGGAAATACCTGGTCTCGGAAACATAAGACACTCTCAGTTTCTCAGGATGAATT 449

Qy     453 ctgggctttcagttatgatgagatggctaggtttgaccttcctgcagtgataaactttat 512
           |||||||||||||||||||||||||||| |  | |||||| || ||    || |||||  ||
```

```
Db    450 CTGGGCTTTCAGTTATGATGAGATGGCAAAATATGACCTACCAGCTTCCATTAACTTCAT 509

Qy    513 tttgcagaaaacgggccaggaaaagatctattatgtcggctattcacagggcaccaccat 572

Db    510 TCTGAATAAAACTGGCCAAGAACAAGTGTATTATGTGGGTCATTCTCAAGGCACCACTAT 569

Qy    573 gggctttattgcattttccaccatgccagagctggctcagaaaatcaaaatgtattttgc 632

Db    570 AGGTTTTATAGCATTTTCACAGATCCCTGAGCTGGCTAAAAGGATTAAAATGTTTTTTGC 629

Qy    633 tttagcacccatagccactgttaagcatgcaaaaagccccgggaccaaatttttgttgct 692

Db    630 CCTGGGTCCTGTGGCTTCCGTCGCCTTCTGTACTAGCCCTATGGCCAAATTAGGACGATT 689

Qy    693 gccagatatgatgatcaagggattgtttggcaaaaaagaatttctgtatcagaccagatt 752

Db    690 ACCAGATCATCTCATTAAGGACTTATTTGGAGACAAAGAATTTCTTCCCCAGAGTGCGTT 749

Qy    753 tctcagacaacttgttatttaccctttgtggccaggtgattcttgatcagatttgtagtaa 812

Db    750 TTTGAAGTGGCTGGGTACCCACGTTTGCACTCATGTCATACTGAAGGAGCTCTGTGGAAA 809

Qy    813 tatcatgttacttctgggtggattcaacaccaacaatatgaacatgagccgagcaagtgt 872

Db    810 TCTCTGTTTTCTTCTGTGTGGATTTAATGAGAGAAATTTAAATATGTCTAGAGTGGATGT 869

Qy    873 atatgctgcccacactcttgctggaacatctgtgcaaaatattctacactggagccaggc 932

Db    870 ATATACAACACATTCTCCTGCTGGAACTTCTGTGCAAAACATGTTACACTGGAGCCAGGC 929

Qy    933 agtgaattctggtgaactccgggcatttgactgggggagtgagaccaaaaatctggaaaa 992

Db    930 TGTTAAATTCCAAAAGTTTCAAGCCTTTGACTGGGGAAGCAGTGCCAAGAATTATTTTCA 989

Qy    993 atgcaatcagccaactcctgtaaggtacagagtcagagatatgacggtccctacagcaat 1052

Db    990 TTACAACCAGAGTTATCCTCCCACATACAATGTGAAGGACATGCTTGTGCCGACTGCAGT 1049

Qy   1053 gtggacaggaggtcaggactggctttcaaatccagaagacgtgaaaatgctgctctctga 1112

Db   1050 CTGGAGCGGGGGTCACGACTGGCTTGCAGATGTCTACGACGTCAATATCTTACTGACTCA 1109

Qy   1113 ggtgaccaacctcatctaccataagaatattcctgaatgggctcacgtggatttcatctg 1172

Db   1110 GATCACCAACTTGGTGTTCCATGAGAGCATTCCGGAATGGGAGCATCTTGACTTCATTTG 1169

Qy   1173 gggtttggatgctcctcaccgtatgtacaatgaaatcatccatctgatgcagcaggagga 1232

Db   1170 GGGCCTGGATGCCCCTTGGAGGCTTTATAATAAAATTATTAATCTAATGAGGAAATATCA 1229

Qy   1233 gaccaacct 1241

Db   1230 GTGAAAGCT 1238
```

```
RESULT   3
HSLAL
LOCUS       HSLAL        2626 bp    mRNA            PRI       25-FEB-1994
DEFINITION  H.sapiens mRNA for lysosomal acid lipase.
ACCESSION   X76488
VERSION     X76488.1  GI:434305
KEYWORDS    lysosomal acid lipase.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2626)
  AUTHORS   Ameis,D., Merkel,M., Eckerskorn,C. and Greten,H.
  TITLE     Purification, characterization and molecular cloning of human
            hepatic lysosomal acid lipase
  JOURNAL   Eur. J. Biochem. 219 (3), 905-914 (1994)
  MEDLINE   94155897
REFERENCE   2  (bases 1 to 2626)
  AUTHORS   Ameis,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-NOV-1993) D. Ameis, Medical Department, University
            Hospital Eppendorf, Martinistrasse 52, 20246 Hamburg, FRG
FEATURES             Location/Qualifiers
     source          1. .2626
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /tissue_type="liver"
     gene            146. .1345
                     /gene="LAL/LIPA"
     CDS             146. .1345
                     /gene="LAL/LIPA"
                     /EC_number="3.1.1.13"
                     /note="sterol esterase"
                     /codon_start=1
                     /evidence=experimental
                     /product="lysosomal acid lipase"
                     /protein_id="CAA54026.1"
                     /db_xref="GI:434306"
                     /db_xref="SWISS-PROT:P38571"
                     /translation="MKMRFLGLVVCLVLWTLHSEGSGGKLTAVDPETNMNVSEIISYW
                     GFPSEEYLVETEDGYILCLNRIPHGRKNHSDKGPKPVVFLQHGLLADSSNWVTNLANS
                     SLGFILADAGFDVWMGNSRGNTWSRKHKTLSVSQDEFWAFSYDEMAKYDLPASINFIL
                     NKTGQEQVYYVGHSQGTTIGFIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGR
                     LPDHLIKDLFGDKEFLPQSAFLKWLGTHVCTHVILKELCGNLCFLLCGFNERNLNMSR
                     VDVYTTHSPAGTSVQNMLHWSQAVKFQKFQAFDWGSSAKNYFHYNQSYPPTYNVKDML
                     VPTAVWSGGHDWLADVYDVNILLTQITNLVFHESIPEWEHLDFIWGLDAPWRLYNKII
                     NLMRKYQ"
     sig_peptide     146. .226
                     /gene="LAL/LIPA"
                     /evidence=experimental
     misc_feature    227. .373
                     /gene="LAL/LIPA"
                     /note="propeptide"
                     /evidence=experimental
     mat_peptide     374. .1342
                     /gene="LAL/LIPA"
                     /evidence=experimental
                     /product="lysosomal acid lipase"
BASE COUNT      741 a    527 c    547 g    811 t
ORIGIN
```

```
  Query Match             41.6%;  Score 527.8;  DB 93;  Length 2626;
  Best Local Similarity   65.4%;  Pred. No. 4.6e-146;
  Matches  791;  Conservative    0;  Mismatches  412;  Indels    6;  Gaps    1;

Qy     39 cagaatggaaatgtggcttctgattctggtggcgtatatgttccagagaaatgtgaattc 98

Db    142 CAGAATGAAAATGCGGTTCTTGGGGTTGGTGGTCTGTTTGGTTCTCTGGACCCTGCATTC 201

Qy     99 ag------tacatatgccaactaaagctgtggacccagaagcattcatgaatattagtga 152

Db    202 TGAGGGTCTGGAGGGAAACTGACAGCTGTGGATCCTGAAACAAACATGAATGTGAGTGA 261

Qy    153 aatcatccaacatcaaggctatccctgtgaggaatatgaagtcgcaactgaagatgggta 212

Db    262 AATTATCTCTTACTGGGGATTCCCTAGTGAGGAATACCTAGTTGAGACAGAAGATGGATA 321

Qy    213 tatcctttctgttaacaggattcctcgaggcctagtgcaacctaagaagacaggttccag 272

Db    322 TATTCTGTGCCTTAACCGAATTCCTCATGGGAGGAAGAACCATTCTGACAAAGGTCCCAA 381

Qy    273 gcctgtggtgttactgcagcatggcctagttggaggtgctagcaactggatttccaacct 332

Db    382 ACCAGTTGTCTTCCTGCAACATGGCTTGCTGGCAGATTCTAGTAACTGGGTCACAAACCT 441

Qy    333 gcccaacaatagcctgggcttcattctggcagatgctggttttgacgtgtggatggggaa 392

Db    442 TGCCAACAGCAGCCTGGGCTTCATTCTTGCTGATGCTGGTTTTGACGTGTGGATGGGCAA 501
```

```
Qy     393 cagcaggggaaacgcctggtctcgaaaacacaagacactctccatagaccaagatgagtt 452
           |||||| |||||  ||||||||| ||||| ||||||||||  |     || ||||| ||
Db     502 CAGCAGAGGAAATACCTGGTCTCGGAAACATAAGACACTCTCAGTTTCTCAGGATGAATT 561

Qy     453 ctgggctttcagttatgatgagatggctaggtttgaccttcctgcagtgataaactttat 512
           ||||||||||||||||||||||||| |  | |||||| || ||    || ||||| ||
Db     562 CTGGGCTTTCAGTTATGATGAGATGGCAAAATATGACCTACCAGCTTCCATTAACTTCAT 621

Qy     513 tttgcagaaaacgggccaggaaaagatctattatgtcggctattcacagggcaccaccat 572
           | || | |||| ||||| ||| |  | ||||||| ||  |||| || ||||||| ||
Db     622 TCTGAATAAAACTGGCCAAGAACAAGTGTATTATGTGGGTCATTCTCAAGGCACCACTAT 681

Qy     573 gggctttattgcattttccaccatgccagagctggctcagaaaatcaaaatgtattttgc 632
            || ||||| |||||||      || || ||||||||| | |  || |||||| ||||||
Db     682 AGGTTTTATAGCATTTTCACAGATCCCTGAGCTGGCTAAAAGGATTAAAATGTTTTTTGC 741

Qy     633 tttagcacccatagccactgttaagcatgcaaaaagccccgggaccaaattttgttgct 692
            | |  || | ||  | ||           |  |||||   |||||||         |
Db     742 CCTGGGTCCTGTGGCTTCCGTCGCCTTCTGTACTAGCCCTATGGCCAAATTAGGACGATT 801

Qy     693 gccagatatgatgatcaagggattgtttggcaaaaaagaatttctgtatcagaccagatt 752
            ||||||     || || ||||  || |||||   | ||||||||||     ||||     ||
Db     802 ACCAGATCATCTCATTAAGGACTTATTTGGAGACAAAGAATTTCTTCCCCAGAGTGCGTT 861

Qy     753 tctcagacaacttgttatttacctttgtggccaggtgattcttgatcagatttgtagtaa 812
           | | |        | | ||  || ||||    || || || ||  |   | || | ||| ||
Db     862 TTTGAAGTGGCTGGGTACCCACGTTTGCACTCATGTCATACTGAAGGAGCTCTGTGGAAA 921

Qy     813 tatcatgttacttctgggtggattcaacaccaacaatatgaacatgagccgagcaagtgt 872
           |. ||    || || |||||| ||||||| ||      |  ||| | || |||     |||     |||
Db     922 TCTCTGTTTTCTTCTGTGTGGATTTAATGAGAGAAATTTAAATATGTCTAGAGTGGATGT 981

Qy     873 atatgctgcccacactcttgctggaacatctgtgcaaaatattctacactggagccaggc 932
           |||| |  | ||   ||| |||||||| |||||||||| ||   |||||||||||||||
Db     982 ATATACAACACATTCTCCTGCTGGAACTTCTGTGCAAAACATGTTACACTGGAGCCAGGC 1041

Qy     933 agtgaattctggtgaactccgggcatttgactgggggagtgagaccaaaaatctggaaaa 992
            || || |       |  | |  || |||||||||| ||      |||| |||       |
Db    1042 TGTTAAATTCCAAAAGTTTCAAGCCTTTGACTGGGGAAGCAGTGCCAAGAATTATTTTCA 1101

Qy     993 atgcaatcagccaactcctgtaaggtacagagtcagagatatgacggtccctacagcaat 1052
            | ||| |||       ||||   |  ||||  || |  || |||  || || || |||· |
Db    1102 TTACAACCAGAGTTATCCTCCCACATACAATGTGAAGGACATGCTTGTGCCGACTGCAGT 1161

Qy    1053 gtggacaggaggtcaggactggctttcaaatccagaagacgtgaaaatgctgctctctga 1112
            ||||  || ||||| |||||||||| || ||  ||||||| |||  || ||  || ||
Db    1162 CTGGAGCGGGGGTCACGACTGGCTTGCAGATGTCTACGACGTCAATATCTTACTGACTCA 1221

Qy    1113 ggtgaccaacctcatctaccataagaatattcctgaatgggctcacgtggatttcatctg 1172
            | | |||||| |  | | ||||| |||  ||||| |||||||  ||  | || ||||| ||
Db    1222 GATCACCAACTTGGTGTTCCATGAGAGCATTCCGGAATGGGAGCATCTTGACTTCATTTG 1281

Qy    1173 gggtttggatgctcctcaccgtatgtacaatgaaatcatccatctgatgcagcaggagga 1232
           |||  ||||||| |||      |  | || ||| |||| ||  |||| |||  | |  |  |
Db    1282 GGGCCTGGATGCCCCTTGGAGGCTTTATAATAAAATTATTAATCTAATGAGGAAATATCA 1341

Qy    1233 gaccaacct 1241
            |  || ||
Db    1342 GTGAAAGCT 1350
```

```
RESULT    4
HUMLIPCHL
LOCUS       HUMLIPCHL     2493 bp    mRNA            PRI       07-JAN-1995
DEFINITION  Human lysosomal acid lipase/cholesteryl esterase mRNA, complete
            cds.
ACCESSION   M74775
VERSION     M74775.1  GI:187151
KEYWORDS    lysosomal acid lipase/cholesteryl esterase.
SOURCE      Homo sapiens fibroblast cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2493)
  AUTHORS   Anderson,R.A. and Sando,G.N.
  TITLE     Cloning and expression of cDNA encoding human lysosomal acid
            lipase/cholesteryl ester hydrolase. Similarities to gastric and
            lingual lipases
  JOURNAL   J. Biol. Chem. 266 (33), 22479-22484 (1991)
  MEDLINE   92042192
FEATURES             Location/Qualifiers
     source          1..2493
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="fibroblast"
                     /map="10"
     mRNA            <1..2493
                     /gene="LIPA"
                     /note="G00-120-153"
     gene            1..2493
                     /gene="LIPA"
     CDS             41..1240
                     /gene="LIPA"
                     /EC_number="3.1.1.13"
                     /note="alternate initiator ATG may be at position 47"
                     /codon_start=1
                     /db_xref="GDB:G00-120-153"
                     /product="lysosomal acid lipase/cholesteryl esterase"
                     /protein_id="AAA59519.1"
                     /db_xref="GI:187152"
                     /translation="MKMRFLGLVVCLVLWPLHSEGSGGKLTAVDPETNMNVSEIISYW
                     GFPSEEYLVETEDGYILCLNRIPHGRKNHSDKGPKPVVFLQHGLLADSSNWVTNLANS
                     SLGFILADAGFDVWMGNSRGNTWSRKHKTLSVSQDEFWAFSYDEMAKYDLPASINFIL
                     NKTGQEQVYYVGHSQGTTIGFIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGR
                     LPDHLIKDLFGDKEFLPQSAFLKWLGTHVCTHVILKELCGNLCFLLCGFNERNLNMSR
                     VDVYTTHSPAGTSVQNMLHWSQAVKFQKFQAFDWGSSAKNYFHYNQSYPPTYNVKDML
                     VPTAVWSGGHDWLADVYDVNILLTQITNLVFHESIPEWEHLDFIWGLDAPWRLYNKII
                     NLMRKYQ"
     sig_peptide     41..103
                     /gene="LIPA"
                     /note="alternate signal peptide may begin at position 47;
                     as predicted from von Heijne, G. (1986) Nucl. Acids Res.
                     14,4683-4690; G00-120-153"
     mat_peptide     104..1237
                     /gene="LIPA"
                     /EC_number="3.1.1.13"
                     /note="G00-120-153"
                     /product="lysosomal acid lipase/cholesteryl esterase"
     polyA_signal    2466..2471
                     /gene="LIPA"
                     /note="G00-120-153"
     polyA_site      2493
                     /gene="LIPA"
                     /note="G00-120-153"
BASE COUNT      692 a     484 c     516 g     801 t
ORIGIN

  Query Match            41.5%;  Score 526.2;  DB 97;  Length 2493;
  Best Local Similarity  65.3%;  Pred. No. 1.4e-145;
  Matches  790;  Conservative    0;  Mismatches  413;  Indels    6;  Gaps    1;
```

```
Qy      39 cagaatggaaatgtggcttctgattctggtggcgtatatgttccagagaaatgtgaattc 98
           ||||||| ||||| || |  ||     ||||||  | | || | |    |    || ||||
Db      37 CAGAATGAAAATGCGGTTCTTGGGGTTGGTGGTCTGTTTGGTTCTCTGGCCCCTGCATTC 96

Qy      99 ag------tacatatgccaactaaagctgtggacccagaagcattcatga
                   |  |   | ||||||||| || ||| ||  |||||
Db      97 TGAGGGGTCTGGAGGGAAACTGACAGCTGTGGATCCTGAAACAAACATGA

Qy     153 aatcatccaacatcaaggctatccctgtgaggaatatgaagtcgcaactga
           ||| |||      |    || |  ||  ||||||||||  |||  |  || ||
Db     157 AATTATCTCTTACTGGGGATTCCCTAGTGAGGAATACCTAGTTGAGACAGA

Qy     213 tatcctttctgttaacaggattcctcgaggcctagtgcaacctaagaagaca
           ||| || |   ||||| | ||||||| ||      | | | |     | |
```

```
Db    217 TATTCTGTGCCTTAACCGAATTCCTCATGGGAGGAAGAACCATTCTGACAAAGGTCCCAA 276

Qy    273 gcctgtggtgttactgcagcatggcctagttggaggtgctagcaactggatttccaacct 332

Db    277 ACCAGTTGTCTTCCTGCAACATGGCTTGCTGGCAGATTCTAGTAACTGGGTCACAAACCT 336

Qy    333 gcccaacaatagcctgggcttcattctggcagatgctggttttgacgtgtggatggggaa 392

Db    337 TGCCAACAGCAGCCTGGGCTTCATTCTTGCTGATGCTGGTTTTGACGTGTGGATGGGCAA 396

Qy    393 cagcaggggaaacgcctggtctcgaaaacacaagacactctccatagaccaagatgagtt 452

Db    397 CAGCAGAGGAAATACCTGGTCTCGGAAACATAAGACACTCTCAGTTTCTCAGGATGAATT 456

Qy    453 ctgggctttcagttatgatgagatggctaggtttgaccttcctgcagtgataaactttat 512

Db    457 CTGGGCTTTCAGTTATGATGAGATGGCAAAATATGACCTACCAGCTTCCATTAACTTCAT 516

Qy    513 tttgcagaaaacgggccaggaaaagatctattatgtcggctattcacagggcaccaccat 572

Db    517 TCTGAATAAAACTGGCCAAGAACAAGTGTATTATGTGGGTCATTCTCAAGGCACCACTAT 576

Qy    573 gggctttattgcattttccaccatgccagagctggctcagaaaatcaaaatgtattttgc 632

Db    577 AGGTTTTATAGCATTTTCACAGATCCCTGAGCTGGCTAAAAGGATTAAAATGTTTTTTGC 636

Qy    633 tttagcacccatagccactgttaagcatgcaaaaagccccgggaccaaatttttgttgct 692

Db    637 CCTGGGTCCTGTGGCTTCCGTCGCCTTCTGTACTAGCCCTATGGCCAAATTAGGACGATT 696

Qy    693 gccagatatgatgatcaagggattgtttggcaaaaaagaatttctgtatcagaccagatt 752

Db    697 ACCAGATCATCTCATTAAGGACTTATTTGGAGACAAAGAATTTCTTCCCCAGAGTGCGTT 756

Qy    753 tctcagacaacttgttatttacctttgtggccaggtgattcttgatcagatttgtagtaa 812

Db    757 TTTGAAGTGGCTGGGTACCCACGTTTGCACTCATGTCATACTGAAGGAGCTCTGTGGAAA 816

Qy    813 tatcatgttacttctgggtggattcaacaccaacaatatgaacatgagccgagcaagtgt 872

Db    817 TCTCTGTTTTCTTCTGTGTGGATTTAATGAGAGAAATTTAAATATGTCTAGAGTGGATGT 876

Qy    873 atatgctgcccacactcttgctggaacatctgtgcaaaatattctacactggagccaggc 932

Db    877 ATATACAACACATTCTCCTGCTGGAACTTCTGTGCAAAACATGTTACACTGGAGCCAGGC 936

Qy    933 agtgaattctggtgaactccgggcatttgactgggggagtgagaccaaaaatctggaaaa 992

Db    937 TGTTAAATTCCAAAAGTTTCAAGCCTTTGACTGGGGAAGCAGTGCCAAGAATTATTTTCA 996

Qy    993 atgcaatcagccaactcctgtaaggtacagagtcagagatatgacggtccctacagcaat 1052

Db    997 TTACAACCAGAGTTATCCTCCCACATACAATGTGAAGGACATGCTTGTGCCGACTGCAGT 1056

Qy   1053 gtggacaggaggtcaggactggctttcaaatccagaagacgtgaaaatgctgctctctga 1112

Db   1057 CTGGAGCGGGGGTCACGACTGGCTTGCAGATGTCTACGACGTCAATATCTTACTGACTCA 1116

Qy   1113 ggtgaccaacctcatctaccataagaatattcctgaatgggctcacgtggatttcatctg 1172

Db   1117 GATCACCAACTTGGTGTTCCATGAGAGCATTCCGGAATGGGAGCATCTTGACTTCATTTG 1176

Qy   1173 gggtttggatgctcctcaccgtatgtacaatgaaatcatccatctgatgcagcaggagga 1232

Db   1177 GGGCCTGGATGCCCCTTGGAGGCTTTATAATAAAATTATTAATCTAATGAGGAAATATCA 1236

Qy   1233 gaccaacct 1241

Db   1237 GTGAAAGCT 1245

RESULT   5
A26689
```

```
LOCUS       A26689       1378 bp    DNA             PAT       05-APR-1995
DEFINITION  Precursor of rabit gastric lipase coding sequence.
ACCESSION   A26689
VERSION     A26689.1  GI:905029
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 1378)
  AUTHORS   Benicourt,C., Blanchard,C. and Junien,J.L.
  TITLE     Recombinant gastric lipase from rabbit and pharmaceutical
            compositions
  JOURNAL   Patent: EP 0542629-A 9 19-MAY-1993;
            INSTITUT DE RECHERCHE JOUVEINAL
FEATURES             Location/Qualifiers
     source          1. .1378
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
BASE COUNT      397 a    287 c    276 g    418 t
ORIGIN
```

Query Match 38.6%; Score 489.2; DB 9; Length 1378;
Best Local Similarity 65.0%; Pred. No. 1.4e-134;
Matches 722; Conservative 0; Mismatches 388; Indels 0; Gaps 0;

```
Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184

Db    136 ACCCTGAAGTGAATATGAATATTAGTCAGATGATTTCCTACTGGGGATACCCAAGTGAAA 195

Qy    185 aatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctcgaggcc 244

Db    196 AATATGAAGTTGTGACTGAAGATGGTTATATCCTTGAGGTTAACAGAATTCCTTATGGGA 255

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304

Db    256 AGAAAAATTCAGGGAACAGAGGCCAGAGACCCGTCGTATTTTTGCAGCATGGTTTGCTTG 315

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364

Db    316 CATCAGCCTCAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTTATTCTGGCAG 375

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424

Db    376 ATGCTGGTTATGGTGTGTGGCTGGGGAACAGCAGAGGAAATACCTGGTCCAGAAGAAACT 435

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484

Db    436 TATACTATTCACCAGACTCCGTTGAATTCTGGGCTTTTAGCTTTGATGAAATGGCTAAAT 495

Qy    485 ttgacattcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 544

Db    496 ATGACCTTCCAGCCACAATTGACTTCATTGTAAAGGAAACTGGACAGGAAAAGCTGCACT 555

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604

Db    556 ATGTTGGTCATTCCCAGGGCACCACCATTGGTTTTATTGCCTTTTCTACAAATCCCAAGC 615

Qy    605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664

Db    616 TGGCTGAAAGAATCAAAACCTTTTATGCATTAGCTCCAGTCGCCACTGTGAAGTATACAA 675

Qy    665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724

Db    676 AAAGCCTTGTAAACAAACTTAGGTTTATTCCTCCAACCATGTTCAAGATTATATTTGGTG 735

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccttttgtggcc 784

Db    736 ACAAAATATTCTACCCACACAATTTCTTTGATCAATTTCTTGCCACTCAAGTGTGTTCCC 795

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844

Db    796 GTGAGACACTGAATGTAATTTGCAGCAATGCCTTATTTATCATTTGTGGCTTTGACAGCG 855
```

```
Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
Db    856 CAAACTTGAACATGAGTCGCTTGGATGTGTACGTGTCACATAATCCGGCAGGAACTTCAG 915

Qy    905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
Db    916 TTCAAAACATGCTGCACTGGACCCAGGCTGTTAAATCTGGAAATTTTCAAGCTTTTAATT 975

Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
Db    976 GGGGAAGTCCAGCTCAGAATGTAGTGCACTTCAATCAGCCCACACCTCCCTACTACAATG 1035

Qy   1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
Db   1036 TGACCGCCATGAATGTGCCAATTGCAGTGTGGAGTGGTGGCAATGACTGGTTGGCTGACC 1095

Qy   1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
Db   1096 CCCAAGATGTTGACCTTTTGCTTCCAAAACTTTCTAATCTCATTTACCACAAGGAGATTC 1155

Qy   1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
Db   1156 TTCCATATAATCACTTGGATTTTATCTGGGCAATGAATGCTCCTCAAGAAGTTTACAATG 1215

Qy   1205 aaatcatccatctgatgcagcaggaggaga 1234
Db   1216 AAATTATTTCTATGATGGCAAAAGATAAAA 1245
```

```
RESULT   6
A26690/c
LOCUS       A26690       1378 bp    DNA              PAT       05-APR-1995
DEFINITION  Precursor of rabit gastric lipase (comp.).
ACCESSION   A26690
VERSION     A26690.1  GI:905030
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 1378)
  AUTHORS   Benicourt,C., Blanchard,C. and Junien,J.L.
  TITLE     Recombinant gastric lipase from rabbit and pharmaceutical
            compositions
  JOURNAL   Patent: EP 0542629-A 10 19-MAY-1993;
            INSTITUT DE RECHERCHE JOUVEINAL
FEATURES             Location/Qualifiers
     source          1..1378
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
BASE COUNT      418 a    276 c    287 g    397 t
ORIGIN

  Query Match            38.6%;  Score 489.2;  DB 9;  Length 1378;
  Best Local Similarity  65.0%;  Pred. No. 1.4e-134;
  Matches  722;  Conservative    0;  Mismatches  388;  Indels    0;  Gaps    0;
```

```
Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
Db   1243 ACCCTGAAGTGAATATGAATATTAGTCAGATGATTTCCTACTGGGGATACCCAAGTGAAA 1184

Qy    185 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 244
Db   1183 AATATGAAGTTGTGACTGAAGATGGTTATATCCTTGAGGTTAACAGAATTCCTTATGGGA 1124

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
Db   1123 AGAAAAATTCAGGGAACAGAGGCCAGAGACCCGTCGTATTTTTGCAGCATGGTTTGCTTG 1064

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
Db   1063 CATCAGCCTCAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTTATTCTGGCAG 1004

Qy    365 atgctggtttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
Db   1003 ATGCTGGTTATGGTGTGTGGCTGGGGAACAGCAGAGGAAATACCTGGTCCAGAAGAAACT 944

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
Db    943 TATACTATTCACCAGACTCCGTTGAATTCTGGGCTTTTAGCTTTGATGAAATGGCTAAAT 884

Qy    485 ttgaccttcctgcagtgataaactttatttgcagaaaacgggccaggaaaagatctatt 544
Db    883 ATGACCTTCCAGCCACAATTGACTTCATTGTAAAGGAAACTGGACAGGAAAAGCTGCACT 824

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
Db    823 ATGTTGGTCATTCCCAGGGCACCACCATTGGTTTTATTGCCTTTTCTACAAATCCCAAGC 764

Qy    605 tggctcagaaaatcaaaatgtatttgctttagcacccatagccactgttaagcatgcaa 664
Db    763 TGGCTGAAAGAATCAAAACCTTTTATGCATTAGCTCCAGTCGCCACTGTGAAGTATACAA 704

Qy    665 aaagccccgggaccaaattttgttgctgccagatatgatgatcaagggattgtttggca 724
Db    703 AAAGCCTTGTAAACAAACTTAGGTTTATTCCTCCAACCATGTTCAAGATTATATTTGGTG 644

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
Db    643 ACAAAATATTCTACCCACACAATTTCTTTGATCAATTTCTTGCCACTCAAGTGTGTTCCC 584

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
Db    583 GTGAGACACTGAATGTAATTTGCAGCAATGCCTTATTTATCATTTGTGGCTTTGACAGCG 524

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
Db    523 CAAACTTGAACATGAGTCGCTTGGATGTGTACGTGTCACATAATCCGGCAGGAACTTCAG 464

Qy    905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
Db    463 TTCAAAACATGCTGCACTGGACCCAGGCTGTTAAATCTGGAAATTTTCAAGCTTTTAATT 404

Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
Db    403 GGGGAAGTCCAGCTCAGAATGTAGTGCACTTCAATCAGCCCACACCTCCCTACTACAATG 344

Qy   1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
Db    343 TGACCGCCATGAATGTGCCAATTGCAGTGTGGAGTGGTGGCAATGACTGGTTGGCTGACC 284

Qy   1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
Db    283 CCCAAGATGTTGACCTTTTGCTTCCAAAACTTTCTAATCTCATTTACCACAAGGAGATTC 224

Qy   1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
Db    223 TTCCATATAATCACTTGGATTTTATCTGGGCAATGAATGCTCCTCAAGAAGTTTACAATG 164

Qy   1205 aaatcatccatctgatgcagcaggaggaga 1234
Db    163 AAATTATTTCTATGATGGCAAAAGATAAAA 134
```

```
RESULT   7
HSGLR
LOCUS       HSGLR        1365 bp    mRNA             PRI       30-MAR-1995
DEFINITION  Human mRNA for gastric lipase.
ACCESSION   X05997
VERSION     X05997.1  GI:31771
KEYWORDS    gastric lipase; lipase.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 1365)
  AUTHORS   Bodmer,M.W., Angal,S., Yarranton,G.T., Harris,T.J., Lyons,A.,
            King,D.J., Pieroni,G., Riviere,C., Verger,R. and Lowe,P.A.
  TITLE     Molecular cloning of a human gastric lipase and expression of the
            enzyme in yeast
  JOURNAL   Biochim. Biophys. Acta 909 (3), 237-244 (1987)
  MEDLINE   87299724
COMMENT     Data kindly reviewed (10-DEC-1987) by LOWE P.A.
FEATURES             Location/Qualifiers
     source          1. .1365
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     sig_peptide     45. .101
                     /product="pot.signal peptide (AA -19 to -1)"
     CDS             45. .1241
                     /codon_start=1
                     /product="gastric lipase precursor"
                     /protein_id="CAA29413.1"
                     /db_xref="GI:758063"
                     /db_xref="SWISS-PROT:P07098"
                     /translation="MWLLLTMASLISVLGTTHGLFGKLHPGSPEVTMNISQMITYWGY
                     PNEEYEVVTEDGYILEVNRIPYGKKNSGNTGQRPVVFLQHGLLASATNWISNLPNNSL
                     AFILADAGYDVWLGNSRGNTWARRNLYYSPDSVEFWAFSFDEMAKYDLPATIDFIVKK
                     TGQKQLHYVGHSQGTTIGFIAFSTNPSLAKRIKTFYALAPVATVKYTKSLINKLRFVP
                     QSLFKFIFGDKIFYPHNFFDQFLATEVCSREMLNLLCSNALFIICGFDSKNFNTSRLD
                     VYLSHNPAGTSVQNMFHWTQAVKSGKFQAYDWGSPVQNRMHYDQSQPPYYNVTAMNVP
                     IAVWNGGKDLLADPQDVGLLLPKLPNLIYHKEIPFYNHLDFIWAMDAPQEVYNDIVSM
                     ISEDKK"
     CDS             63. .1241
                     /codon_start=1
                     /product="gastric lipase precursor"
                     /protein_id="CAA29414.1"
                     /db_xref="GI:758064"
                     /db_xref="SWISS-PROT:P07098"
                     /translation="MASLISVLGTTHGLFGKLHPGSPEVTMNISQMITYWGYPNEEYE
                     VVTEDGYILEVNRIPYGKKNSGNTGQRPVVFLQHGLLASATNWISNLPNNSLAFILAD
                     AGYDVWLGNSRGNTWARRNLYYSPDSVEFWAFSFDEMAKYDLPATIDFIVKKTGQKQL
                     HYVGHSQGTTIGFIAFSTNPSLAKRIKTFYALAPVATVKYTKSLINKLRFVPQSLFKF
                     IFGDKIFYPHNFFDQFLATEVCSREMLNLLCSNALFIICGFDSKNFNTSRLDVYLSHN
                     PAGTSVQNMFHWTQAVKSGKFQAYDWGSPVQNRMHYDQSQPPYYNVTAMNVPIAVWNG
                     GKDLLADPQDVGLLLPKLPNLIYHKEIPFYNHLDFIWAMDAPQEVYNDIVSMISEDKK
                     "
     sig_peptide     63. .101
                     /product="pot.signal peptide (AA -13 to -1)"
     mat_peptide     102. .1238
                     /product="gastric lipase (AA 1-379)"
BASE COUNT      396 a    284 c    268 g    417 t
ORIGIN
```

Query Match 37.0%; Score 470; DB 93; Length 1365;
Best Local Similarity 64.2%; Pred. No. 7.3e-129;
Matches 707; Conservative 0; Mismatches 395; Indels 0; Gaps 0;

```
Qy    119 ctgtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccct 178
Db    121 CTGGAAGCCCTGAAGTGACTATGAACATTAGTCAGATGATTACTTATTGGGGATACCCAA 180

Qy    179 gtgaggaatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctc 238
Db    181 ATGAAGAATATGAAGTTGTGACTGAAGATGGTTATATTCTTGAAGTCAATAGAATTCCTT 240

Qy    239 gaggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcc 298
Db    241 ATGGGAAGAAAAATTCAGGGAATACAGGCCAGAGACCTGTTGTGTTTTTGCAGCATGGTT 300

Qy    299 tagttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattc 358
Db    301 TGCTTGCATCAGCCACAAACTGGATTTCCAACCTGCCGAACAACAGCCTTGCCTTCATTC 360

Qy    359 tggcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaa 418
Db    361 TGGCAGATGCTGGTTATGATGTGTGGCTGGGCAACAGCAGAGGAAACACCTGGGCCAGAA 420

Qy    419 aacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatgg 478
Db    421 GAAACTTGTACTATTCACCAGATTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGG 480

Qy    479 ctaggtttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaaga 538
Db    481 CTAAATATGACCTTCCAGCCACAATCGACTTCATTGTAAAGAAAACTGGACAGAAGCAGC 540

Qy    539 tctattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgc 598
Db    541 TACACTATGTTGGCCATTCCCAGGGCACCACCATTGGTTTTATTGCCTTTTCCACCAATC 600

Qy    599 cagagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagc 658
Db    601 CCAGCCTGGCTAAAAGAATCAAAACCTTCTATGCTCTAGCTCCTGTTGCCACTGTGAAGT 660

Qy    659 atgcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgt 718
Db    661 ATACAAAAAGCCTTATAAACAAACTTAGATTTGTTCCTCAATCCCTCTTCAAGTTTATAT 720

Qy    719 ttggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccttt 778
Db    721 TTGGTGACAAAATATTCTACCCACACAACTTCTTTGATCAATTTCTTGCTACTGAAGTGT 780

Qy    779 gtggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattca 838
Db    781 GCTCCCGTGAGATGCTGAATCTCCTTTGCAGCAATGCCTTATTTATAATTTGTGGATTTG 840

Qy    839 acaccaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaa 898
Db    841 ACAGTAAGAACTTTAACACGAGTCGCTTGGATGTGTATCTATCACATAATCCAGCAGGAA 900

Qy    899 catctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcat 958
Db    901 CTTCTGTTCAAAACATGTTCCATTGGACCCAGGCTGTTAAGTCTGGGAAATTCCAAGCTT 960

Qy    959 ttgactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggt 1018
Db    961 ATGACTGGGGAAGCCCAGTTCAGAATAGGATGCACTATGATCAGTCCCAACCTCCCTACT 1020

Qy   1019 acagagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggcttt 1078
Db   1021 ACAATGTGACAGCCATGAATGTACCAATTGCAGTGTGGAACGGTGGCAAGGACCTGTTGG 1080

Qy   1079 caaatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataaga 1138
Db   1081 CTGACCCCCAAGATGTTGGCCTTTTGCTTCCAAAACTCCCCAATCTTATTTACCACAAGG 1140

Qy   1139 atattcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgt 1198
Db   1141 AGATTCCTTTTTACAATCACTTGGACTTTATCTGGGCAATGGATGCCCCTCAAGAAGTTT 1200

Qy   1199 acaatgaaatcatccatctgat 1220
Db   1201 ACAATGACATTGTTTCTATGAT 1222
```

```
RESULT   8
A01046
LOCUS       A01046       1367 bp    mRNA            PAT       16-MAR-1993
DEFINITION  H.sapiens mRNA for human gastric lipase.
ACCESSION   A01046
VERSION     A01046.1  GI:344241
KEYWORDS    gastric lipase.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1367)
  AUTHORS   .
  TITLE     POLYPEPTIDE AND POLYPEPTIDE COMPOSITION
  JOURNAL   Patent: WO 8601532-A 5 13-MAR-1986;
```

```
FEATURES             Location/Qualifiers
     source          1. .1367
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     sig_peptide     47. .103
     CDS             47. .1243
                     /note="gastric lipase"
                     /codon_start=1
                     /product="lipase"
                     /protein_id="CAA00125.1"
                     /db_xref="GI:344242"
                     /db_xref="SWISS-PROT:P07098"
                     /translation="MWLLLTMASLISVLGTTHGLFGKLHPGSPEVTMNISQMITYWGY
                     PNEEYEVVTEDGYILEVNRIPYGKKNSGNTGQRPVVFLQHGLLASATNWISNLPNNSL
                     AFILADAGYDVWLGNSRGNTWARRNLYYSPDSVEFWAFSFDEMAKYDLPATIDFIVKK
                     TGQKQLHYVGHSQGTTIGFIAFSTNPSLAKRIKTFYALAPVATVKYTKSLINKLRFVP
                     QSLFKFIFGDKIFYPHNFFDQFLATEVCSREMLNLLCSNALFIICGFDSKNFNTSRLD
                     VYLSHNPAGTSVQNMFHWTQAVKSGKFQAYDWGSPVQNRMHYDQSQPPYYNVTAMNVP
                     IAVWNGGKDLLADPQDVGLLLPKLPNLIYHKEIPFYNHLDFIWAMDAPQEVYNDIVSM
                     ISEDKK"
     mat_peptide     104. .1240
                     /note="gastric lipase"
                     /product="lipase"
BASE COUNT      397 a    284 c    269 g    417 t
ORIGIN
```

```
  Query Match             37.0%;  Score 470;  DB 9;  Length 1367;
  Best Local Similarity   64.2%;  Pred. No. 7.3e-129;
  Matches  707;  Conservative    0;  Mismatches  395;  Indels    0;  Gaps    0;

Qy    119 ctgtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccct 178
          |||    ||| ||||     ||||| |||||| | || ||      ||   || || ||
Db    123 CTGGAAGCCCTGAAGTGACTATGAACATTAGTCAGATGATTACTTATTGGGGATACCCAA 182

Qy    179 gtgaggaatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctc 238
          ||| ||||||||||| |  ||||||||||| ||||| |||   || || || ||||||
Db    183 ATGAAGAATATGAAGTTGTGACTGAAGATGGTTATATTCTTGAAGTCAATAGAATTCCTT 242

Qy    239 gaggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcc 298
          ||        |  |   ||| |||||    || ||||| |||||  ||||||||||
Db    243 ATGGGAAGAAAAATTCAGGGAATACAGGCCAGAGACCTGTTGTGTTTTTGCAGCATGGTT 302

Qy    299 tagttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattc 358
          |  ||| |   || |  |||||||||||||||||| ||||| ||||| | |||||||||
Db    303 TGCTTGCATCAGCCACAAACTGGATTTCCAACCTGCCGAACAACAGCCTTGCCTTCATTC 362

Qy    359 tggcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaa 418
          ||||||||||||||| ||| |||||| |||| |||||||| |||||| |||||  |  |||
Db    363 TGGCAGATGCTGGTTATGATGTGTGGCTGGGCAACAGCAGAGGAAACACCTGGGCCAGAA 422

Qy    419 aacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatgg 478
          | ||  |      ||   |||   || ||| |||||||||||||| | ||||||  ||||
Db    423 GAAACTTGTACTATTCACCAGATTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGG 482

Qy    479 ctaggtttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaaaga 538
          |||  | |||||||| ||     ||  |||| ||| |  ||||||| || ||| |  ||
Db    483 CTAAATATGACCTTCCAGCCACAATCGACTTCATTGTAAAGAAAACTGGACAGAAGCAGC 542

Qy    539 tctattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgc 598
          |  | ||||| ||| |||| |||||||||||||| || |||||||| ||||||||||  |
Db    543 TACACTATGTTGGCCATTCCCAGGGCACCACCATTGGTTTTATTGCCTTTTCCACCAATC 602

Qy    599 cagagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagc 658
          |    |||||| | | ||||||||  |  | |||| |||| ||  | ||||||||| |||
Db    603 CCAGCCTGGCTAAAAGAATCAAAACCTTCTATGCTCTAGCTCCTGTTGCCACTGTGAAGT 662

Qy    659 atgcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgt 718
          ||·|||||||||     | |||| ||   || | || |     | |||||    | |
Db    663 ATACAAAAAGCCTTATAAACAAACTTAGATTTGTTCCTCAATCCCTCTTCAAGTTTATAT 722

Qy    719 ttggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttt 778
          ||||  | |||  |||       || | |  |||     |   |||| || | |  | |
```

```
Db    723 TTGGTGACAAAATATTCTACCCACACAACTTCTTTGATCAATTTCTTGCTACTGAAGTGT 782

Qy    779 gtggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattca 838
          |   ||  | ||| ||  |||   |||| || |||  | | ||  |  | |||||||
Db    783 GCTCCCGTGAGATGCTGAATCTCCTTTGCAGCAATGCCTTATTTATAATTTGTGGATTTG 842

Qy    839 acaccaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaa 898
          ||| .|| ||  | |||| ||| ||      ||| |||    | || | ||  || ||||
Db    843 ACACTAAGAACTTTAACACGAGTCGCTTGGATGTGTATCTATCACATAATCCAGCAGGAA 902

Qy    899 catctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcat 958
          | ||||| ||||| ||  | || |||| |||||| || || ||||  || |||  || |
Db    903 CTTCTGTTCAAAACATGTTCCATTGGACCCAGGCTGTTAAGTCTGGGAAATTCCAAGCTT 962

Qy    959 ttgactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggt 1018
          |||||||||  ||        | |||  |    | |   ||||| |    |||      |
Db    963 ATGACTGGGGAAGCCCAGTTCAGAATAGGATGCACTATGATCAGTCCCAACCTCCCTACT 1022

Qy    1019 acagagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggcttt 1078
           |||  || | ||  ||||  || || |  ||| ||||||  || ||  |||||  | |
Db    1023 ACAATGTGACAGCCATGAATGTACCAATTGCAGTGTGGAACGGTGGCAAGGACCTGTTGG 1082

Qy    1079 caaatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataaga 1138
           |  | ||  |||| ||      |  |||  |  |  |  |||| || || ||||| |||
Db    1083 CTGACCCCCAAGATGTTGGCCTTTTGCTTCCAAAACTCCCCAATCTTATTTACCACAAGG 1142

Qy    1139 atattcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgt 1198
           | ||||||   |    |||| |||| || |||||||   ||||||| |||||     | |
Db    1143 AGATTCCTTTTTACAATCACTTGGACTTTATCTGGGCAATGGATGCCCCTCAAGAAGTTT 1202

Qy    1199 acaatgaaatcatccatctgat 1220
           ||||||| ||  |   | ||||
Db    1203 ACAATGACATTGTTTCTATGAT 1224
```

```
RESULT   9
A12714
LOCUS       A12714         1367 bp    DNA              PAT       09-AUG-1994
DEFINITION  Pregastric lipase.
ACCESSION   A12714
VERSION     A12714.1  GI:579560
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1367)
  AUTHORS   .
  TITLE     PRECURSOR POLYPEPTIDE, DNA SEQUENCE CODING THEREFOR, VECTORS, HOST
            ORGANISMS, AND PROCESSES INVOLVING SAME
  JOURNAL   Patent: WO 8603778-A 8 03-JUL-1986;
FEATURES             Location/Qualifiers
     source          1. .1367
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             47. .1243
                     /codon_start=1
                     /product="pregastric lipase"
                     /protein_id="CAA01053.1"
                     /db_xref="GI:579561"
                     /db_xref="SWISS-PROT:P07098"
                     /translation="MWLLLTMASLISVLGTTHGLFGKLHPGSPEVTMNISQMITYWGY
                     PNEEYEVVTEDGYILEVNRIPYGKKNSGNTGQRPVVFLQHGLLASATNWISNLPNNSL
                     AFILADAGYDVWLGNSRGNTWARRNLYYSPDSVEFWAFSFDEMAKYDLPATIDFIVKK
                     TGQKQLHYVGHSQGTTIGFIAFSTNPSLAKRIKTFYALAPVATVKYTKSLINKLRFVP
                     QSLFKFIFGDKIFYPHNFFDQFLATEVCSREMLNLLCSNALFIICGFDSKNFNTSRLD
                     VYLSHNPAGTSVQNMFHWTQAVKSGKFQAYDWGSPVQNRMHYDQSQPPYYNVTAMNVP
                     IAVWNGGKDLLADPQDVGLLLPKLPNLIYHKEIPFYNHLDFIWAMDAPQEVYNDIVSM
                     ISEDKK"
BASE COUNT      397 a    284 c    269 g    417 t
ORIGIN
```

```
  Query Match           37.0%;  Score 470;  DB 9;  Length 1367;
  Best Local Similarity  64.2%;  Pred. No. 7.3e-129;
  Matches  707;  Conservative    0;  Mismatches  395;  Indels    0;  Gaps    0;

Qy    119 ctgtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccct 178
Db    123 CTGGAAGCCCTGAAGTGACTATGAACATTAGTCAGATGATTACTTATTGGGGATACCCAA 182

Qy    179 gtgaggaatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctc 238
Db    183 ATGAAGAATATGAAGTTGTGACTGAAGATGGTTATATTCTTGAAGTCAATAGAATTCCTT 242

Qy    239 gaggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcc 298
Db    243 ATGGGAAGAAAAATTCAGGGAATACAGGCCAGAGACCTGTTGTGTTTTTGCAGCATGGTT 302

Qy    299 tagttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattc 358
Db    303 TGCTTGCATCAGCCACAAACTGGATTTCCAACCTGCCGAACAACAGCCTTGCCTTCATTC 362

Qy    359 tggcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaa 418
Db    363 TGGCAGATGCTGGTTATGATGTGTGGCTGGGCAACAGCAGAGGAAACACCTGGGCCAGAA 422

Qy    419 aacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatgg 478
Db    423 GAAACTTGTACTATTCACCAGATTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGG 482

Qy    479 ctaggtttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaaga 538
Db    483 CTAAATATGACCTTCCAGCCACAATCGACTTCATTGTAAAGAAAACTGGACAGAAGCAGC 542

Qy    539 tctattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgc 598
Db    543 TACACTATGTTGGCCATTCCCAGGGCACCACCATTGGTTTTATTGCCTTTTCCACCAATC 602

Qy    599 cagagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagc 658
Db    603 CCAGCCTGGCTAAAAGAATCAAAACCTTCTATGCTCTAGCTCCTGTTGCCACTGTGAAGT 662

Qy    659 atgcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgt 718
Db    663 ATACAAAAAGCCTTATAAACAAACTTAGATTTGTTCCTCAATCCCTCTTCAAGTTTATAT 722

Qy    719 ttggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccttt 778
Db    723 TTGGTGACAAAATATTCTACCCACACAACTTCTTTGATCAATTTCTTGCTACTGAAGTGT 782

Qy    779 gtggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattca 838
Db    783 GCTCCCGTGAGATGCTGAATCTCCTTTGCAGCAATGCCTTATTTATAATTTGTGGATTTG 842

Qy    839 acaccaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaa 898
Db    843 ACAGTAAGAACTTTAACACGAGTCGCTTGGATGTGTATCTATCACATAATCCAGCAGGAA 902

Qy    899 catctgtgcaaaatattctacactggagccaggcagtgaatttctggtgaactccgggcat 958
Db    903 CTTCTGTTCAAAACATGTTCCATTGGACCCAGGCTGTTAAGTCTGGGAAATTCCAAGCTT 962

Qy    959 ttgactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggt 1018
Db    963 ATGACTGGGGAAGCCCAGTTCAGAATAGGATGCACTATGATCAGTCCCAACCTCCCTACT 1022

Qy   1019 acagagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggcttt 1078
Db   1023 ACAATGTGACAGCCATGAATGTACCAATTGCAGTGTGGAACGGTGGCAAGGACCTGTTGG 1082

Qy   1079 caaatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataaga 1138
Db   1083 CTGACCCCCAAGATGTTGGCCTTTTGCTTCCAAAACTCCCCAATCTTATTTACCACAAGG 1142

Qy   1139 atattcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgt 1198
Db   1143 AGATTCCTTTTTACAATCACTTGGACTTTATCTGGGCAATGGATGCCCCTCAAGAAGTTT 1202

Qy   1199 acaatgaaatcatccatctgat 1220
Db   1203 ACAATGACATTGTTTCTATGAT 1224


RESULT  10
A01157
LOCUS       A01157       1336 bp    mRNA            PAT       29-JAN-1993
DEFINITION  R.norvegicus mRNA for prelingual lipase protein.
ACCESSION   A01157
VERSION     A01157.1  GI:14748
KEYWORDS    lipase.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1336)
  AUTHORS   Williamson,R.
  JOURNAL   Patent: WO 8500381-A 3 31-JAN-1985;
            CellTech Ltd
FEATURES             Location/Qualifiers
     source          1. .1336
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
     CDS             7. .1194
                     /codon_start=1
                     /product="prelingual lipase"
                     /protein_id="CAA00136.1"
                     /db_xref="GI:14749"
                     /db_xref="SWISS-PROT:P04634"
                     /translation="MWLLLITSVISTFGGAHGLFGKLGPGNPEANMNISQMITYWGYP
                     CQEYEVVTEDGYILGVYRIPHGKNNSENIGKRPVVYLQHGLIASATNWIANLPNNSLA
                     FMLADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATINFIVQKT
                     GQEKIHYVGHSQGTTIGFIAFSTNPTLAKKIKTFYALAPVATVKYTQSPLKKISFIPT
                     FLFKLMFGKKMFLPHTYFDDFLGTEVCSREVLDLLCSNTLFIFCGFDKKNLNVSRFDV
                     YLGHNPAGTSVQDFLHWAQLVRSGKFQAFNWGSPSQNMLHYNQKTPPEYDVSAMTVPV
                     AVWNGGNDILADPQDVAMLLPKLSNLLFHKEILAYNHLDFIWAMDAPQEVYNEMISMM
                     AED"
BASE COUNT      409 a    283 c    265 g    379 t
ORIGIN

  Query Match           36.8%;  Score 467.6;  DB 9;  Length 1336;
  Best Local Similarity  63.5%;  Pred. No. 3.8e-128;
  Matches  713;  Conservative    0;  Mismatches  409;  Indels    0;  Gaps    0;

Qy    119 ctgtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccct 178
Db     80 CTGGAAACCCTGAAGCAAATATGAATATTAGTCAGATGATAACTTACTGGGGATATCCAT 139

Qy    179 gtgaggaatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctc 238
Db    140 GTCAAGAATATGAAGTTGTTACTGAAGATGGCTACATTCTGGGGGTCTACAGAATTCCTC 199

Qy    239 gaggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcc 298
Db    200 ATGGGAAGAATAATTCTGAAAATATAGGCAAGAGACCTGTGGTGTATTTGCAGCATGGTT 259

Qy    299 tagttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattc 358
Db    260 TGATCGCATCAGCCACAAACTGGATTGCAAATCTACCAAACAACAGCCTGGCCTTTATGC 319

Qy    359 tggcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaa 418
Db    320 TAGCAGACGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGAAATACATGGTCCAGGA 379

Qy    419 aacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatgg 478
Db    380 AAAATGTATACTACTCACCAGACTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGG 439
```

```
Qy     479 ctaggtttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaaga 538
Db     440 CTAAATATGACCTTCCCGCCACAATAAACTTCATTGTACAGAAAACTGGACAAGAGAAGA 499

Qy     539 tctattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgc 598
Db     500 TACACTATGTTGGTCATTCTCAGGGCACCACTATTGGTTTCATTGCCTTTTCTACCAATC 559

Qy     599 cagagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagc 658
Db     560 CTACACTGGCCAAAAAAATCAAGACGTTTTATGCATTAGCTCCAGTTGCTACCGTGAAGT 619

Qy     659 atgcaaaaagccccgggaccaaattttttgttgctgccagatatgatgatcaagggattgt 718
Db     620 ATACACAAAGTCCCTTGAAAAAGATTTCATTTATTCCTACATTTCTTTTCAAGCTTATGT 679

Qy     719 ttggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccttt 778
Db     680 TTGGCAAGAAAATGTTCCTGCCCCACACCTACTTTGATGACTTTCTTGGTACCGAAGTGT 739

Qy     779 gtggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattca 838
Db     740 GCTCACGGGAGGTTCTAGATCTTCTCTGCAGCAACACTTTATTCATCTTCTGTGGATTTG 799

Qy     839 acaccaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaa 898
Db     800 ACAAGAAAAACTTAAATGTGAGTCGTTTTGATGTGTATCTAGGGCATAATCCAGCAGGGA 859

Qy     899 catctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcat 958
Db     860 CATCTGTTCAAGACTTTCTCCACTGGGCACAGCTTGTTAGATCTGGGAAATTTCAAGCCT 919

Qy     959 ttgactggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggt 1018
Db     920 TCAACTGGGGAAGCCCATCCCAGAACATGTTACACTACAACCAGAAAACGCCTCCTGAAT 979

Qy    1019 acagagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggcttt 1078
Db     980 ATGATGTGTCAGCCATGACTGTGCCAGTTGCAGTGTGGAACGGTGGCAATGACATCCTGG 1039

Qy    1079 caaatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataaga 1138
Db    1040 CTGATCCCCAAGATGTCGCCATGCTGCTTCCCAAACTCTCCAACCTCCTGTTCCATAAGG 1099

Qy    1139 atattcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgt 1198
Db    1100 AGATTCTTGCCTACAATCACCTGGACTTCATCTGGGCAATGGATGCCCCTCAAGAGGTTT 1159

Qy    1199 acaatgaaatcatccatctgatgcagcaggaggagaccaacc 1240
Db    1160 ACAATGAGATGATTTCCATGATGGCAGAAGACTAAAAGAATC 1201
```

```
RESULT  11
RNLIP
LOCUS       RNLIP        1355 bp    mRNA             ROD       30-MAR-1995
DEFINITION  Rat mRNA for lingual lipase.
ACCESSION   X02309
VERSION     X02309.1  GI:56595
KEYWORDS    glycoprotein; lipase.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1355)
  AUTHORS   Docherty,A.J., Bodmer,M.W., Angal,S., Verger,R., Riviere,C.,
            Lowe,P.A., Lyons,A., Emtage,J.S. and Harris,T.J.
  TITLE     Molecular cloning and nucleotide sequence of rat lingual lipase
            cDNA
  JOURNAL   Nucleic Acids Res. 13 (6), 1891-1903 (1985)
  MEDLINE   85215587
COMMENT     Data kindly reviewed (12-FEB-1986) by A.J.P. Docherty.
FEATURES             Location/Qualifiers
     source          1..1355
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
     sig_peptide     49..102
                     /product="put. signal peptide (aa -18 to -1)"
     CDS             49..1236
                     /note="lipase precursor"
                     /codon_start=1
                     /protein_id="CAA26179.1"
                     /db_xref="GI:56596"
                     /db_xref="SWISS-PROT:P04634"
                     /translation="MWLLLITSVISTFGGAHGLFGKLGPGNPEANMNISQMITYWGYP
                     CQEYEVVTEDGYILGVYRIPHGKNNSENIGKRPVVYLQHGLIASATNWIANLPNNSLA
                     FMLADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATINFIVQKT
                     GQEKIHYVGHSQGTTIGFIAFSTNPTLAKKIKTFYALAPVATVKYTQSPLKKISFIPT
                     FLFKLMFGKKMFLPHTYFDDFLGTEVCSREVLDLLCSNTLFIFCGFDKKNLNVSRFDV
                     YLGHNPAGTSVQDFLHWAQLVRSGKFQAFNWGSPSQNMLHYNQKTPPEYDVSAMTVPV
                     AVWNGGNDILADPQDVAMLLPKLSNLLFHKEILAYNHLDFIWAMDAPQEVYNEMISMM
                     AED"
     mat_peptide     103..1233
                     /product="lipase (aa 1-377)"
     misc_feature    505..507
                     /note="aa 153 pot. lipid binding site"
     misc_feature    1355
                     /note="polyA site"
BASE COUNT      400 a    291 c    276 g    387 t      1 others
ORIGIN

  Query Match            36.8%;  Score 467.6;  DB 95;  Length 1355;
  Best Local Similarity   63.5%;  Pred. No. 3.8e-128;
  Matches  713;  Conservative    0;  Mismatches  409;  Indels    0;  Gaps    0;

Qy     119 ctgtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccct 178
Db     122 CTGGAAACCCTGAAGCAAATATGAATATTAGTCAGATGATAACTTACTGGGGATATCCAT 181

Qy     179 gtgaggaatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctc 238
Db     182 GTCAAGAATATGAAGTTGTTACTGAAGATGGCTACATTCTGGGGGTCTACAGAATTCCTC 241

Qy     239 gaggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcc 298
Db     242 ATGGGAAGAATAATTCTGAAAATATAGGCAAGAGACCTGTGGTGTATTTGCAGCATGGTT 301

Qy     299 tagttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattc 358
Db     302 TGATCGCATCAGCCACAAACTGGATTGCAAATCTACCAAACAACAGCCTGGCCTTTATGC 361

Qy     359 tggcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaa 418
Db     362 TAGCAGACGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGAAATACATGGTCCAGGA 421

Qy     419 aacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatgg 478
Db     422 AAAATGTATACTACTCACCAGACTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGG 481

Qy     479 ctacgtttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaaga 538
Db     482 CTAAATATGACCTTCCCGCCACAATAAACTTCATTGTACAGAAAACTGGACAAGAGAAGA 541

Qy     539 tctattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgc 598
Db     542 TACACTATGTTGGTCATTCTCAGGGCACCACTATTGGTTTCATTGCCTTTTCTACCAATC 601

Qy     599 cagagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagc 658
Db     602 CTACACTGGCCAAAAAAATCAAGACGTTTTATGCATTAGCTCCAGTTGCTACCGTGAAGT 661

Qy     659 atgcaaaaagccccgggaccaaattttttgttgctgccagatatgatgatcaagggattgt 718
```

```
Db    662 ATACACAAAGTCCCTTGAAAAAGATTTCATTTATTCCTACATTTCTTTTCAAGCTTATGT 721

Qy    719 ttggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccttt 778
          ||||||| |||   || |||  || |||  |||          |||| ||   | | |
Db    722 TTGGCAAGAAAATGTTCCTGCCCCACACCTACTTTGATGACTTTCTTGGTACCGAAGTGT 781

Qy    779 gtggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattca 838
          |    | || | |||| ||||   | || || || | |  | ||  |  |  |||||||
Db    782 GCTCACGGGAGGTTCTAGATCTTCTCTGCAGCAACACTTTATTCATCTTCTGTGGATTTG 841

Qy    839 acaccaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaa 898
          |||  || ||  | ||  |||| ||      ||| |||   |  || | ||  || || |
Db    842 ACAAGAAAAACTTAAATGTGAGTCGTTTTGATGTGTATCTAGGGCATAATCCAGCAGGGA 901

Qy    899 catctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcat 958
          ||||||| ||| |  |||| |||||   |||   || |  |||||  || | |  || |
Db    902 CATCTGTTCAAGACTTTCTCCACTGGGCACAGCTTGTTAGATCTGGGAAATTTCAAGCCT 961

Qy    959 ttgactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggt 1018
          |  ||||||| ||     || | ||  ||  | | | ||| |||  ||| |||      |
Db    962 TCAACTGGGGAAGCCCATCCCAGAACATGTTACACTACAACCAGAAAACGCCTCCTGAAT 1021

Qy   1019 acagagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggcttt 1078
          |    ||   ||  ||||| || ||     ||| ||||||  || ||  | |||   ||
Db   1022 ATGATGTGTCAGCCATGACTGTGCCAGTTGCAGTGTGGAACGGTGGCAATGACATCCTGG 1081

Qy   1079 caaatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataaga 1138
          |  ||||  |||| ||    ||||||||  |  |  | ||||||| | | ||||||||
Db   1082 CTGATCCCCAAGATGTCGCCATGCTGCTTCCCAAACTCTCCAACCTCCTGTTCCATAAGG 1141

Qy   1139 atattcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgt 1198
          | |||| ||  |    |||| |||| ||||||||||   ||||||| |||||     | |
Db   1142 AGATTCTTGCCTACAATCACCTGGACTTCATCTGGGCAATGGATGCCCCTCAAGAGGTTT 1201

Qy   1199 acaatgaaatcatccatctgatgcagcaggaggagaccaacc 1240
          ||||||| || ||     |||||    | ||  | |  || |
Db   1202 ACAATGAGATGATTTCCATGATGGCAGAAGACTAAAAGAATC 1243


RESULT  12
MMLYACLY
LOCUS       MMLYACLY     2358 bp    mRNA            ROD       24-FEB-1999
DEFINITION  M.musculus (C57 Black/6X CBA) LAL mRNA for lysosomal acid lipase.
ACCESSION   Z31689
VERSION     Z31689.1  GI:4456670
KEYWORDS    LAL; lipase; lysosomal acid lipase.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2358)
  AUTHORS   Du,H., Witte,D.P. and Grabowski,G.A.
  TITLE     Tissue and cellular specific expression of murine lysosomal acid
            lipase mRNA and protein
  JOURNAL   J. Lipid Res. 37 (5), 937-949 (1996)
  MEDLINE   96363957
REFERENCE   2  (bases 1 to 2358)
  AUTHORS   Du,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-APR-1994) Hong Du, Division of Human Genetics,
            Children's Hospital Medical Center, 3333 Bernet Street, Cincinnati,
            OH, 45229-3039, USA
FEATURES             Location/Qualifiers
     source          1. .2358
                     /organism="Mus musculus"
                     /strain="C57 Black/6X CBA"
                     /db_xref="taxon:10090"
                     /sex="Female"
                     /clone="mLAL #49"
                     /tissue_type="liver"
                     /clone_lib="Mouse liver cDNA library /Lamda-ZAP"
     gene            19. .1212
                     /gene="LAL"
     CDS             19. .1212
                     /gene="LAL"
                     /citation=[1]
                     /codon_start=1
                     /product="lysosomal acid lipase"
                     /protein_id="CAA83494.1"
                     /db_xref="GI:4456671"
                     /translation="MQLQGLVFVFTIGILLSRVPTGTVSAVDPEVNMNVTEIIMRWGY
                     PGEEHSVLTGDGYILSIHRIPRGWKNHFGKGPRPVVYLQHGLLADSSNWVTNIDNSSL
                     GFLLADRGFDVWMGNSRGNTWSLKHKTLSVSQDEFWAFSFDEMAKYDLPASINYILNK
                     TGQEQIYYVGHSQGCTIGFIAFSQMPELAKKIKMFLVLAPVLSLNFASGPLLQLGRLP
                     DPLLKDMFGQKQFLPQSAMLKWLSIHVCTHVIMKELCANVFFLLCGFNEKNLNMSRVD
                     VYTTHCPAELLVQNMLHWGQVFKYRKLQAFDWGSSEKNYFHYNQSFPPSYNIKNMRLP
                     TALWSGGRDWLADINDITILLTQIPKLVYHKNIPEWDHLDFIWGLDAPWKLYDEIISL
                     MKKYQ"
     sig_peptide     19. .75
                     /gene="LAL"
                     /product="lysosomal acid lipase"
     mat_peptide     76. .1209
                     /gene="LAL"
                     /product="lysosomal acid lipase"
     polyA_signal    2246. .2251
     polyA_site      2358
BASE COUNT      604 a    549 c    524 g    681 t
ORIGIN


  Query Match           36.3%;  Score 461;  DB 94;  Length 2358;
  Best Local Similarity   63.3%;  Pred. No. 3.8e-126;
  Matches  707;  Conservative    0;  Mismatches  410;  Indels    0;  Gaps    0;

Qy    117 agctgtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatcc 176
          |||||||| || ||||    |||||| | |  || || |||   |     || |||||
Db     93 AGCTGTGGATCCTGAAGTCAACATGAATGTGACCGAGATAATCATGCGCTGGGGATATCC 152

Qy    177 ctgtgaggaatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcc 236
          |||||||| |    |||  || | ||||| ||||| ||   | || || | || ||
Db    153 TGGTGAGGAACACTCGGTCCTGACAGGAGATGGCTATATTCTGAGTATTCACCGAATCCC 212

Qy    237 tcgaggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatgg 296
          ||| ||      | | | |      | |||| |||| || || |||  || ||||| ||
Db    213 TCGTGGGTGGAAGAACCATTTTGGCAAAGGTCCCAGACCAGTTGTGTATCTTCAGCACGG 272

Qy    297 cctagttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcat 356
          | |  | | || | |||| |||||  |  ||| |   |||||  ||| |||||||| |
Db    273 CTTGCTGGCAGATTCTAGTAACTGGGTCACAAACATTGACAACAGCAGCTTGGGCTTCCT 332

Qy    357 tctggcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcg 416
          |||||  |||  ||| ||||| |||||||||| |||||||| |||||| |||||||||
Db    333 CCTGGCGGATCGTGGCTTTGATGTGTGGATGGGAAACAGCAGAGGAAACACCTGGTCTCT 392

Qy    417 aaaacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagat 476
          || |||||||| || ||  |    || ||||||||||||||| |||||| ||||||||
Db    393 GAAGCACAAGACTCTTTCAGTTTCTCAGGATGAGTTGTGGGCCTTCAGTTTTGATGAGAT 452

Qy    477 ggctaggtttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaa 536
          ||| |  | |||||| |||||    || |||| |||||| | ||||| ||||| ||| |
Db    453 GGCAAAATATGACCTACCTGCTTCAATTAACTACATTTTGAATAAAACTGGCCAAGAACA 512

Qy    537 gatctattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccat 596
          ||||||||||| ||  |||| || |||  |||||| || || || || |||||    ||
Db    513 AATCTATTATGTGGGTCATTCTCAAGGCTGCACCATAGGTTTCATAGCTTTTTCACAGAT 572

Qy    597 gccagagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaa 656
          ||| ||| ||||  |||| || ||||||| ||| |  || || ||| |     || | ||
Db    573 GCCTGAGTTGGCAAAGAAGATTAAAATGTTTTTAGTCTTGGCTCCCGTGTTGTCTCTCAA 632

Qy    657 gcatgcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggatt 716
          |||   | ||||        ||||       |||||||     | ||||||  |
Db    633 TTTTGCTTCAGGCCCGCTACTTCAATTGGGACGCTTGCCAGACCCTCTTCTCAAGGACAT 692

Qy    717 gtttggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacct 776
          ||||||  | ||  |||||||  |||||    | || |     ||    ||| || |
```

```
Db    693 GTTTGGACAGAAGCAATTTCTCCCTCAGAGTGCGATGCTGAAGTGGCTGAGCATTCACGT 752

Qy    777 ttgtggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggatt 836
          |||   ||| || ||  |  |  || ||||    ||  || | || || ||  |||| ||
Db    753 TTGCACCCACGTCATCATGAAGGAGCTTTGCGCAAACGTCTTCTTCCTGCTATGTGGCTT 812

Qy    837 caacaccaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctgg 896
          |||    || ||| | || |||    |||    ||| ||| |  | |||   | ||| |
Db    813 CAATGAGAAGAATTTAAATATGTCTAGAGTGGATGTGTATACAACACACTGCCCTGCGGA 872

Qy    897 aacatctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggc 956
          |       |||||||| ||  |||||||| | ||||   | || |   || | || |  ||
Db    873 ACTTCTCGTGCAAAACATGTTACACTGGGGTCAGGTTTTTAAATACCGTAAGCTTCAAGC 932

Qy    957 atttgactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaag 1016
           |||||||||||| ||        || |||       | | ||| |||     ||||
Db    933 TTTTGACTGGGGAAGCAGTGAAAAGAATTATTTTCATTACAACCAGAGTTTTCCTCCCTC 992

Qy   1017 gtacagagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggct 1076
          |||||   | | | | |||   | || || ||  ||||||  || || | |||||||||
Db    993 GTACAACATAAAAAACATGCGTCTACCCACTGCTTTGTGGAGCGGGGGCCGGGACTGGCT 1052

Qy   1077 ttcaaatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataa 1136
           || |     | ||| | |  ||  | ||  || | |  |||| |  | || || ||
Db   1053 GGCAGACATCAATGACATCACTATCTTACTGACTCAAATCCCCAAGTTGGTGTATCACAA 1112

Qy   1137 gaatattcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtat 1196
          ||| |||||||| |||| ||  ||||||| ||||||||||||||||| ||        |
Db   1113 GAACATTCCTGAGTGGGACCATCTGGATTTTATCTGGGGTTTGGATGCCCCCTGGAAACT 1172

Qy   1197 gtacaatgaaatcatccatctgatgcagcaggaggag 1233
          |||  ||||||| |||  ||| ||| || |  |  ||
Db   1173 GTATGATGAAATAATCAGTCTAATGAAGAAATACCAG 1209


RESULT  13
S81497
LOCUS       S81497       3144 bp    mRNA            ROD       02-AUG-2000
DEFINITION  lysosomal acid lipase=intracellular hydrolase [rats, Wolman, liver,
            mRNA, 3144 nt].
ACCESSION   S81497
VERSION     S81497.1  GI:1336725
KEYWORDS    .
SOURCE      Rattus sp. liver Wolman.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 3144)
  AUTHORS   Nakagawa,H., Matsubara,S., Kuriyama,M., Yoshidome,H., Fujiyama,J.,
            Yoshida,H. and Osame,M.
  TITLE     Cloning of rat lysosomal acid lipase cDNA and identification of the
            mutation in the rat model of Wolman's disease
  JOURNAL   J. Lipid Res. 36 (10), 2212-2218 (1995)
  MEDLINE   96129534
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 176569] from the original journal article.
            This sequence comes from Fig. 2.
FEATURES             Location/Qualifiers
     source          1. .3144
                     /organism="Rattus sp."
                     /db_xref="taxon:10118"
     gene            1. .3144
                     /gene="lysosomal acid lipase, LAL"
     CDS             233. .1426
                     /gene="lysosomal acid lipase, LAL"
                     /note="intracellular hydrolase; This sequence comes from
                     Fig. 2; conceptual translation presented here differs from
                     translation in publication; LAL"
                     /codon_start=1
                     /product="lysosomal acid lipase"
                     /protein_id="AAB36043.2"
                     /db_xref="GI:9653291"
                     /translation="MQLLGRVICFVVGILLSGGPTGTISAVDPEANMNVTEIIMHWGY
                     PEHSVQTGDGYILGVHRIPHGRKNQFDKGPKPVVYLQWRHGFLADSSNWVTNIDNNSL
                     GFILADAGFDVWMGNSRGNTWSRKHKTLSVSQDEYWAFSFDEMAKYDLPASINYILNK
                     TGQEQLYNVGHSQGCTIGFIAFSQMPELAKKVKMFFALAPVLSLNFASGPMVKLGRLP
                     DLLLEDLFGQKQFLPQSAMVKWLSTHICTHVIMKELCANIFFLICGFNEKNLNMSRVD
                     VYTTHCPAGTSVQNMVHWTQVVKYHKLQAFDWGSSDKNYFHYNQSYPPLYSIKDMQLP
                     TALWSGGKDWLADTSDINILLTEIPTLVYHKNIPEWDHLDFIWGLDAPWRLYNEVVSL
                     MKKYQ"
BASE COUNT      852 a    690 c    754 g    848 t
ORIGIN


  Query Match             36.2%;  Score 459.8;  DB 95;  Length 3144;
  Best Local Similarity   64.5%;  Pred. No. 9e-126;
  Matches  724;  Conservative    0;  Mismatches  387;  Indels   12;  Gaps    2;

Qy    117 agctgtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatcc 176
          |||||||||||| ||||||  |||||||  | | ||| || |||   ||    || ||||||
Db    307 AGCTGTGGACCCTGAAGCAAACATGAATGTGACTGAGATAATCATGCACTGGGGATATCC 366

Qy    177 ctgtgaggaatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcc 236
              ||  |   |||    || | |||||| || || ||   |||| || | || ||
Db    367 A------GAGCACTCAGTACAGACAGGAGATGGCTACATTCTGGGTGTTCACCGAATCCC 420

Qy    237 tcgaggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcag----- 291
           |  || |    | | |       | | |||| |||  || || ||||  || |||
Db    421 GCATGGGCGGAAGAACCAGTTTGATAAAGGTCCCAAACCAGTTGTGTATCTTCAGTGGCG 480

Qy    292 -catggcctagttggaggtgctagcaactggatttccaacctgcccaacaatagcctggg 350
           |||||| |  | | || | |||| |||||| |  | ||| |   ||||| ||||||||
Db    481 ACATGGCTTCCTGGCAGATTCTAGTAACTGGGTCACAAACATTGACAACAACAGCCTGGG 540

Qy    351 cttcattctggcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctg 410
          |||||| |||||||||||||| ||||| |||||||||||| |||||||| |||||| ||||
Db    541 CTTCATCCTGGCAGATGCTGGCTTTGATGTGTGGATGGGAAACAGCAGAGGAAACACCTG 600

Qy    411 gtctcgaaaacacaagacactctccatagaccaagatgagttctgggctttcagttatga 470
          ||||||. || |||||||| || ||  |    || ||||| | |||||| ||||||| |||
Db    601 GTCTCGGAAGCACAAGACTCTGTCAGTTTCTCAGGATGAATACTGGGCCTTCAGTTTTGA 660

Qy    471 tgagatggctaggtttgaccttcctgcagtgataaacttttattttgcagaaaacgggcca 530
          |||||||| |  | |||||| ||||    || ||||  |||||| | ||||| |||||
Db    661 TGAGATGGCAAAATATGACCTCCCTGCTTCAATTAACTACATTTTGAATAAAACTGGCCA 720

Qy    531 ggaaaagatctattatgtcggctattcacagggcaccaccatgggctttattgcattttc 590
          ||| || | ||| |||| ||| | || || ||| |||||| || ||||| ||||||||
Db    721 AGAACAGCTGTATAATGTGGGCCACTCTCAAGGCTGCACCATAGGTTTTATAGCATTTTC 780

Qy    591 caccatgccagagctggctcagaaaatcaaaatgtattttgctttagcacccatagccac 650
              ||||| || |||||  ||||  | |||||| ||||||  | || ||  |     |
Db    781 ACAGATGCCTGAACTGGCAAAGAAGGTTAAAATGTTTTTTGCCCTGGCTCCTGTGTTGTC 840

Qy    651 tgttaagcatgcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaa 710
          | |  ||   |||     ||||   | ||||||       | |||| ||| || | |  |
Db    841 TCTCAATTTTGCTTCGGGCCCGATGGTCAAATTGGGACGGTTGCCGGATCTGCTTCTTGA 900

Qy    711 gggattgtttggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttat 770
          ||  |||||||  | ||  |||||||   |||||    | | | | |    ||    |
Db    901 GGACTTGTTTGGGCAGAAGCAATTTCTTCCTCAGAGTGCAATGGTGAAATGGCTAAGCAC 960

Qy    771 ttacctttgtggccaggtgattcttgatcagatttgtagtaatatcatgttacttctggg 830
          | || ||||    || || ||  |  |  || ||||    || ||| | || || |  |
Db    961 TCACATTTGCACGCACGTCATCATGAAGGAGCTTTGCGCCAACATCTTCTTTCTGATATG 1020

Qy    831 tggattcaacaccaacaatatgaacatgagccgagcaagtgtatatgctgcccacactct 890
          ||| |||||    || ||| | ||||||    |||    ||| ||| |  | |||  ||
Db   1021 TGGCTTCAATGAGAAGAATTTAAACATGTCTAGAGTGGATGTGTATACAACACACTGTCC 1080

Qy    891 tgctggaacatctgtgcaaaatattctacactggagccaggcagtgaattctggtgaact 950
          ||| ||||| |||||||||| ||  |||||||| |||||   || || |    | | ||
Db   1081 TGCCGGAACCTCTGTGCAAAACATGGTACACTGGACCCAGGTTGTTAAATACCATAAGCT 1140

Qy    951 ccgggcatttgactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcc 1010
```

```
Db   1141 TCAAGCTTTTGACTGGGGAAGCAGTGACAAGAATTATTTTCATTACAATCAGAGTTATCC 1200

Qy   1011 tgtaaggtacagagtcagagatatgacggtccctacagcaatgtggacaggaggtcagga 1070

Db   1201 ACCCTTGTACAGCATAAAAGACATGCAACTACCCACTGCTTTATGGAGCGGGGGCAAGGA 1260

Qy   1071 ctggctttcaaatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatcta 1130

Db   1261 CTGGCTGGCAGACACCAGTGACATCAATATCTTACTGACTGAAATCCCCACGTTAGTGTA 1320

Qy   1131 ccataagaatattcctgaatgggctcacgtggatttcatctggggtttggatgctcctca 1190

Db   1321 CCACAAGAACATTCCCGAGTGGGACCATCTGGATTTTATCTGGGGTTTGGATGCCCCCTG 1380

Qy   1191 ccgtatgtacaatgaaatcatccatctgatgcagcaggaggag 1233

Db   1381 GAGGCTGTATAATGAAGTAGTCAGTCTAATGAAGAAATACCAG 1423
```

```
RESULT  14
BOVMPE
LOCUS       BOVMPE       1417 bp    mRNA            MAM       10-DEC-1994
DEFINITION  Bos taurus pregastric esterase mRNA, complete cds.
ACCESSION   L26319
VERSION     L26319.1  GI:600756
KEYWORDS    esterase; pregastric esterase.
SOURCE      Bos taurus 25 weeks Tongue serous gland cDNA to mRNA.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 1417)
  AUTHORS   Timmermans,M.Y.J., Kupers,L.P. and Teuchy,H.
  TITLE     The cDNA sequence encoding bovine pregastric esterase
  JOURNAL   Gene 147, 259-262 (1994)
  MEDLINE   95011625
FEATURES             Location/Qualifiers
     source          1. .1417
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /cell_type="Secretory cell"
                     /dev_stage="25 weeks"
                     /tissue_type="Tongue serous gland"
     sig_peptide     73. .129
     CDS             73. .1266
                     /EC_number="3.1.1.3"
                     /codon_start=1
                     /product="pregastric esterase"
                     /protein_id="AAA57037.1"
                     /db_xref="GI:600757"
                     /translation="MWWLLVTVCFIHMSGNAFCFLGKIAKNPEASMNVSQMISYWGYP
                     SEMHKVITADGYILQVYRIPHGKNNANHLGQRPVVFLQHGLLGSATNWISNLPKNSLG
                     FLLADAGYDVWLGNSRGNTWAQEHLYYSPDSPEFWAFSFDEMAEYDLPSTIDFILRRT
                     GQKKLHYVGHSQGTTIGFIAFSTSPTLAEKIKVFYALAPVATVKYTKSLFNKLALIPH
                     FLFKIIFGDKMFYPHTFLEQFLGVEMCSRETLDVLCKNALFAITGVDNKNFNMSRLDV
                     YIAHNPAGTSVQNTLHWRQAVKSGKFQAFDWGAPYQNLMHYHQPTPPIYNLTAMNVPI
                     AVWSADNDLLADPQDVDFLLSKLSNLIYHKEIPNYNHLDFIWAMDAPQEVYNEIVSLM
                     AEDKK"
     mat_peptide     130. .1263
                     /EC_number="3.1.1.3"
                     /function="triacylglycerol acylhydrolase"
                     /evidence=experimental
                     /product="pregastric esterase"
     polyA_signal    1378. .1385
     polyA_site      1396
BASE COUNT      408 a    306 c    276 g    427 t
ORIGIN
```

```
  Query Match           36.1%;  Score 458;  DB 7;  Length 1417;
  Best Local Similarity  62.1%;  Pred. No. 2.8e-125;
  Matches  741;  Conservative    0;  Mismatches  450;  Indels    3;  Gaps    1;
```

```
Qy     44 tggaaatgtggcttctgattctggtggcgtatatgttccagagaaatgtgaattcagtac 103

Db     68 TGAAAATGTGGTGGCTACTTGTAACAGTGTGTTTCATCCACATGTCTGGAAATGCATTTT 127

Qy    104 atatgc---caactaaagctgtggacccagaagcattcatgaatattagtgaaatcatcc 160

Db    128 GTTTCCTTGGAAAAATTGCTAAGAACCCTGAAGCCAGTATGAATGTTAGTCAGATGATTT 187

Qy    161 aacatcaaggctatccctgtgaggaatatgaagtcgcaactgaagatgggtatatccttt 220

Db    188 CCTACTGGGGCTACCCAAGTGAGATGCATAAAGTTATAACTGCGGATGGTTATATCCTTC 247

Qy    221 ctgttaacaggattcctcgaggcctagtgcaacctaagaagacaggttccaggcctgtgg 280

Db    248 AGGTCTATCGGATTCCTCATGGAAAGAATAATGCTAATCATTTAGGTCAGAGACCTGTTG 307

Qy    281 tgttactgcagcatggcctagttggaggtgctagcaactggatttccaacctgcccaaca 340

Db    308 TGTTTCTGCAGCATGGTCTTCTTGGATCAGCCACAAACTGGATTTCCAACCTGCCCAAGA 367

Qy    341 atagcctgggcttcattctggcagatgctggttttgacgtgtggatggggaacagcaggg 400

Db    368 ACAGCCTGGGCTTCCTCCTGGCAGATGCTGGTTATGACGTGTGGCTGGGGAACAGCAGAG 427

Qy    401 gaaacgcctggtctcgaaaacacaagacactctccatagaccaagatgagttctgggctt 460

Db    428 GAAACACCTGGGCCCAGGAACATTTATACTATTCACCAGACTCCCCGGAATTCTGGGCTT 487

Qy    461 tcagttatgatgagatggctaggtttgaccttcctgcagtgataaacttttattttgcaga 520

Db    488 TCAGCTTTGATGAAATGGCGGAATATGACCTTCCATCTACAATTGATTTCATCTTAAGGA 547

Qy    521 aaacgggccaggaaaagatctattatgtcggctattcacagggcaccaccatgggcttta 580

Db    548 GAACAGGACAGAAGAAGCTACACTATGTTGGCCATTCCCAAGGCACCACCATTGGTTTTA 607

Qy    581 ttgcattttccaccatgccagagctggctcagaaaatcaaaatgtattttgctttagcac 640

Db    608 TCGCCTTTTCTACCAGTCCCACATTGGCTGAAAAAATCAAAGTCTTCTATGCATTAGCCC 667

Qy    641 ccatagccactgttaagcatgcaaaaagccccgggaccaaatttttgttgctgccagata 700

Db    668 CAGTTGCCACAGTGAAGTACACCAAGAGCCTGTTTAACAAACTTGCACTTATTCCTCACT 727

Qy    701 tgatgatcaagggattgtttggcaaaaaagaatttctgtatcagaccagatttctcagac 760

Db    728 TCCTCTTCAAGATTATATTTGGTGACAAAATGTTCTACCCACACACTTTTTTGGAACAAT 787

Qy    761 aacttgttatttaccttttgtggccaggtgattcttgatcagatttgtagtaatatcatgt 820

Db    788 TTCTTGGTGTTGAAATGTGCTCCCGTGAGACACTGGATGTCCTTTGTAAGAATGCCTTGT 847

Qy    821 tacttctgggtggattcaacaccaacaatatgaacatgagccgagcaagtgtatatgctg 880

Db    848 TTGCCATTACTGGAGTTGACAATAAAAACTTCAACATGAGTCGCTTAGATGTGTATATAG 907

Qy    881 cccaactcttgctggaacatctgtgcaaaatattctacactggagccaggcagtgaatt 940

Db    908 CACATAATCCAGCAGGAACTTCTGTTCAAAACACCCTCCACTGGAGACAGGCTGTTAAGT 967

Qy    941 ctggtgaactccgggcatttgactgggggagtgagaccaaaaatctggaaaaatgcaatc 1000

Db    968 CTGGGAAATTCCAAGCTTTTGACTGGGGAGCCCCATATCAGAACCTAATGCATTATCATC 1027

Qy   1001 agccaactcctgtaaggtacagagtcagagatatgacggtccctacagcaatgtggacag 1060

Db   1028 AGCCCACACCTCCCATCTACAATTTAACAGCCATGAATGTCCCAATTGCAGTATGGAGTG 1087

Qy   1061 gaggtcaggactggctttcaaatccagaagacgtgaaaatgctgctctctgaggtgacca 1120

Db   1088 CTGACAATGACCTGTTGGCTGACCCTCAGGATGTTGACTTTCTGCTTTCAAAACTCTCTA 1147
```

```
Qy   1121 acctcatctaccataagaatattcctgaatgggctcacgtggatttcatctggggtttgg 1180
          | ||||| ||||| ||| | ||||| | |    |||| |||| || |||||||   |||
Db   1148 ATCTCATTTACCACAAGGAAATTCCAAATTACAATCACTTGGACTTTATCTGGGCAATGG 1207

Qy   1181 atgctcctcaccgtatgtacaatgaaatcatccatctgatgcagcaggaggaga 1234
          |||| |||||     | ||||||||||| |   | |||||    | || | |
Db   1208 ATGCACCTCAAGAAGTTTACAATGAAATTGTTTCTTTGATGGCCGAAGACAAAA 1261


RESULT  15
AR039022
LOCUS       AR039022     1137 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 2 from patent US 5807726.
ACCESSION   AR039022
VERSION     AR039022.1  GI:5958385
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1137)
  AUTHORS   Blanchard,C., Benicourt,C. and Junien,J.
  TITLE     Nucleic acids encoding dog gastric lipase and their use for the
            production of polypeptides
  JOURNAL   Patent: US 5807726-A 2 15-SEP-1998;
FEATURES             Location/Qualifiers
     source          1. .1137
                     /organism="unknown"
BASE COUNT      296 a    308 c    249 g    284 t
ORIGIN


  Query Match           35.0%;  Score 444.2;  DB 9;  Length 1137;
  Best Local Similarity   62.8%;  Pred. No. 3.5e-121;
  Matches  689;  Conservative    0;  Mismatches  408;  Indels    0;  Gaps    0;

Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
          |||| ||||    ||||||||| ||| | || |||    |    || || ||   |||||
Db     26 ACCCTGAAGTGACCATGAATATAAGTCAGATGATCACCTACTGGGGATACCCAGCTGAGG 85

Qy    185 aatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctcgaggcc 244
          ||||||||||  |  || ||||| || |||||||||    |  |||| ||||||   ||
Db     86 AATATGAAGTTGTGACCGAAGACGGTTATATCCTTGGGATCGACAGAATTCCTTATGGGA 145

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
            |  |  |||| | |||    || ||||| |  ||  |||| || ||  |  | |
Db    146 GGAAAAATTCAGAGAATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCG 205

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
            |  || |  |||||||| |||||||||||||||| ||||||| |||||| ||||| |
Db    206 CATCAGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCCTGGCCG 265

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
          | || || |  |||||||| |||||||||||||||| ||| ||||| |  | |   |
Db    266 ACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACCTGGGCCAGGAGGAATC 325

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
            |     |||    |||   |  || ||||||||||||| | |||.||||||||||  |
Db    326 TGTACTACTCGCCCGACTCCGTCGAATTCTGGGCTTTCAGCTTTGACGAGATGGCTAAAT 385

Qy    485 ttgaccttcctgcagtgataaactttatttttgcagaaaacgggccaggaaaagatctatt 544
            |||||||| ||    ||  |||| || ||| |||||||||| ||||| ||| |  | |
Db    386 ATGACCTTCCCGCCACCATTGACTTCATCTTGAAGAAAACGGGACAGGACAAGCTACACT 445

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
          | || ||| |||| |||||||||||||| || || || || ||||||||||  ||  |||
Db    446 ACGTTGGCCATTCCCAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGC 505

Qy    605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
          ||||  |    |||||||  |  | ||| ||||| ||| | ||||| || ||| |  |
Db    506 TGGCGAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGTACACCG 565

Qy    665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
          ||| ||     | |||| |  || |  | ||     |  |  |||||   | |||||  |

Db    566 AAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTCAAGCTTATATTTGGAA 625

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
          | |||  |||       ||   |  |||     |  || |  |  |  | ||   ||
Db    626 ACAAAATATTCTACCCACACCACTTCTTTGATCAATTTCTCGCCACCGAGGTATGCTCCC 685

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
            | ||   | ||||   | || || ||   | ||||  |  |  |||||||  |||| |
Db    686 GCCAGACGGTGGATCTCCTCTGCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTA 745

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
            ||  |||||||||| ||      ||| |||    | || | ||  || |||||||| |
Db    746 TGAACTTGAACATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGG 805

Qy    905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
          | || ||. |.|| |||||| |||||| || || |||||  |  |||  || |||||||
Db    806 TTCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAGCTTTTGACT 865

Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
          |||| ||         | ||  ||    | |   |||||   |  |||      ||||
Db    866 GGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGCATGCCTCCCTACTACAACC 925

Qy   1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
          | | ||| |||   || || |  ||| ||||||  || ||  | |||| |||  |  | |
Db    926 TGACAGACATGCATGTGCCAATCGCAGTGTGGAACGGTGGCAACGACTTGCTGGCCGACC 985

Qy   1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
          |  | || ||  |  | |||| ||  || |  |||| ||||| ||||| | ||| ||||
Db    986 CTCACGATGTTGACCTTTTGCTTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTC 1045

Qy   1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
          ||   |    |||| |||| || |||||||   ||||||| |||||     | |||||||
Db   1046 CTCCTTACAATCACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATG 1105

Qy   1205 aaatcatccatctgatg 1221
          ||||  |     |||||
Db   1106 AAATTGTTTCCATGATG 1122


Search completed: August 24, 2001, 22:44:47
Job time: 9169 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: August 24, 2001, 22:41:42 ; Search time 211.8 Seconds
(without alignments)
3762.073 Million cell updates/sec

Title: US-09-333-159-46
Perfect score: 1269
Sequence: 1 atgttggaaaccttgtcaag.....gacggtgtgaggccgtattg 1269

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601:*
1: /SIDS8/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1269	100.0	1269	22	AAF45132	Human TANGO 294 OR
2	1269	100.0	2044	22	AAF45131	Human TANGO 294 cD
3	489.2	38.6	1378	14	AAQ42310	RGL precursor. Or
4	470	37.0	1367	7	AAN60685	Sequence encoding
5	470	37.0	1367	7	AAN60566	Sequence encoding
6	470	37.0	1367	17	AAT58916	Human gastric lipa
7	464.4	36.6	1336	6	AAN50385	Rat prelingual lip
8	444.2	35.0	1528	15	AAQ68388	Canine gastric lip
9	444.2	35.0	1528	17	AAT58915	Dog gastric lipase
10	444.2	35.0	1531	15	AAQ68389	Canine gastric lip
11	286	22.5	699	22	AAF28679	Human protein HP03
12	285	22.5	1308	22	AAF28689	Human protein HP03
13	257.8	20.3	617	21	AAA44349	Human secreted exp
14	193.2	15.2	572	21	AAA43744	Mouse secreted exp
15	100	7.9	1718	21	AAZ34958	Soybean acid triac
16	89.4	7.0	1483	21	AAZ34956	Rice acid triacylg
17	84.8	6.7	859	21	AAZ34953	Corn acid triacylg
18	72	5.7	1438	21	AAZ34959	Soybean acid triac
c 19	68	5.4	936	22	AAF58252	Oligonucleotide D1
c 20	68	5.4	936	22	AAF58254	Oligonucleotide D1
c 21	68	5.4	936	22	AAF58257	Oligonucleotide D1
c 22	68	5.4	936	22	AAF58259	Oligonucleotide D2
c 23	68	5.4	936	22	AAF58262	Oligonucleotide D2
c 24	68	5.4	938	22	AAF58255	Oligonucleotide D1
25	66.2	5.2	936	22	AAF58252	Oligonucleotide D1
26	66.2	5.2	936	22	AAF58254	Oligonucleotide D1
27	66.2	5.2	936	22	AAF58257	Oligonucleotide D1
28	66.2	5.2	936	22	AAF58259	Oligonucleotide D2
29	66.2	5.2	936	22	AAF58262	Oligonucleotide D2
30	66.2	5.2	938	22	AAF58255	Oligonucleotide D1
31	55.8	4.4	705	21	AAZ34952	Catalpa acid triac
32	54	4.3	549	21	AAC93827	Cat flea hindgut a
33	51	4.0	509	21	AAZ34954	Corn acid triacylg
34	48.2	3.8	178	19	AAV35431	Cell death regulat
35	46.4	3.7	43676	21	AAZ60606	Nucleotide sequenc
c 36	38.6	3.0	1664976	19	AAV21209	Methanococcus jann
c 37	36.8	2.9	13254	21	AAA32165	Complete nucleotid
c 38	36.6	2.9	244	22	AAF58238	Oligonucleotide D1
39	36.6	2.9	3271	19	AAV00060	Human G-protein co
40	36.6	2.9	3282	17	AAT40914	DNA encoding VH re
41	36.6	2.9	3282	21	AAA31024	Modified heavy cha
c 42	36.6	2.9	3282	21	AAA31038	B12 IgG1 preferred
43	36.6	2.9	3282	21	AAA32150	Modified heavy cha
c 44	36.6	2.9	3282	21	AAA32164	B12 IgG1 preferred
45	36.6	2.9	5412	20	AAX21356	Human BAI2 gene.

ALIGNMENTS

RESULT 1
AAF45132
ID AAF45132 standard; cDNA; 1269 BP.
XX
AC AAF45132;
XX
DT 30-MAR-2001 (first entry)
XX
DE Human TANGO 294 ORF.
XX
KW Gene therapy; TANGO protein; INTERCEPT protein; neurological disorder;
KW central nervous system; focal brain disorder; bipolar affective disorder;
KW global-diffuse cerebral disorder; cerebrovascular; Alzheimer's disease;
KW senile dementia; Huntington's disease; amyotrophic lateral sclerosis;
KW Parkinson's; Gilles de la Tourette's syndrome; hypertension; sleep;
KW neuropsychiatric; psychoactive substance use; anxiety; ss.
XX
OS Homo sapiens.
XX
PN WO200077239-A2.
XX
PD 21-DEC-2000.
XX
PF 24-MAY-2000; 2000WO-US14858.
XX
PR 14-JUN-1999; 99US-0333159.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI McCarthy SA, Fraser CC, Sharp JD, Barnes TM;
XX
DR WPI; 2001-032313/04.
DR P-PSDB; AAB66065.
XX

PT TANGO and INTERCEPT nucleic acids, proteins, and antibodies, useful for
PT screening assays and diagnostic assays and for the treatment of
PT neurological diseases such as Alzheimer's, Parkinson's and Huntington's
PT disease
XX
PS Claim 1; Fig 6; 359pp; English.
XX
CC The present invention relates to TANGO or INTERCEPT proteins and coding
CC sequences (see AAF45121-F45136 and AAF45138-F45139 and AAB66031-B66057,
CC AAB66064-B66083 and AAB66085). The TANGO/INTERCEPT proteins and coding
CC sequences are useful for the treatment of neurological disorders such as
CC central nervous system (CNS) disorders, CNS-related disorders, focal
CC brain disorders, global-diffuse cerebral disorders and other
CC neurological and cerebrovascular disorders. The CNS disorders include
CC Alzheimer's disease, senile dementia, Huntington's disease, amyotrophic
CC lateral sclerosis, Parkinson's, Gilles de la Tourette's syndrome,
CC autonomic function disorders such as hypertension and sleep disorders,
CC neuropsychiatric disorders, psychoactive substance use disorders,
CC anxiety, and bipolar affective disorder.
XX
SQ Sequence 1269 BP; 358 A; 268 C; 309 G; 334 T; 0 other;

Query Match 100.0%; Score 1269; DB 22; Length 1269;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 atgttggaaaccttgtcaagacagtggattgtctcacacagaatggaaatgtggcttctg 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 atgttggaaaccttgtcaagacagtggattgtctcacacagaatggaaatgtggcttctg 60

Qy     61 attctggtggcgtatatgttccagagaaatgtgaattcagtacatatgccaactaaagct 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 attctggtggcgtatatgttccagagaaatgtgaattcagtacatatgccaactaaagct 120

Qy    121 gtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgt 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 gtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgt 180

Qy    181 gaggaatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcga 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 gaggaatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcga 240

Qy    241 ggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggccta 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggccta 300

Qy    301 gttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctg 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 gttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctg 360

Qy    361 gcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaa 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 gcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaa 420

Qy    421 cacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatggct 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 cacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatggct 480

Qy    481 aggtttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 aggtttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatc 540

Qy    541 tattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgcca 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 tattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgcca 600

Qy    601 gagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcat 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 gagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcat 660

Qy    661 gcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgttt 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 gcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgttt 720

Qy    721 ggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgt 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 ggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgt 780

Qy    781 ggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaac 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 ggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaac 840

Qy    841 accaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaaca 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 accaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaaca 900

Qy    901 tctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcattt 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 tctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcattt 960

Qy    961 gactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtac 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 gactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtac 1020

Qy   1021 agagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttca 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 agagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttca 1080

Qy   1081 aatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaat 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 aatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaat 1140

Qy   1141 attcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtac 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 attcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtac 1200

Qy   1201 aatgaaatcatccatctgatgcagcaggaggagaccaacctttcccagggacggtgtgag 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 aatgaaatcatccatctgatgcagcaggaggagaccaacctttcccagggacggtgtgag 1260

Qy   1261 gccgtattg 1269
          |||||||||
Db   1261 gccgtattg 1269
```

RESULT 2
AAF45131
ID AAF45131 standard; cDNA; 2044 BP.
XX
AC AAF45131;
XX
DT 30-MAR-2001 (first entry)
XX
DE Human TANGO 294 cDNA.
XX
KW Gene therapy; TANGO protein; INTERCEPT protein; neurological disorder;
KW central nervous system; focal brain disorder; bipolar affective disorder;
KW global-diffuse cerebral disorder; cerebrovascular; Alzheimer's disease;
KW senile dementia; Huntington's disease; amyotrophic lateral sclerosis;
KW Parkinson's; Gilles de la Tourette's syndrome; hypertension; sleep;
KW neuropsychiatric; psychoactive substance use; anxiety; ss.
XX
OS Homo sapiens.
XX
PN WO200077239-A2.
XX
PD 21-DEC-2000.
XX
PF 24-MAY-2000; 2000WO-US14858.
XX
PR 14-JUN-1999; 99US-0333159.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI McCarthy SA, Fraser CC, Sharp JD, Barnes TM;

```
XX
DR   WPI; 2001-032313/04.
DR   P-PSDB; AAB66065.
XX
PT   TANGO and INTERCEPT nucleic acids, proteins, and antibodies, useful for
PT   screening assays and diagnostic assays and for the treatment of
PT   neurological diseases such as Alzheimer's, Parkinson's and Huntington's
PT   disease -
XX
PS   Claim 1; Fig 6; 359pp; English.
XX
CC   The present invention relates to TANGO or INTERCEPT proteins and coding
CC   sequences (see AAF45121-F45136 and AAF45138-F45139 and AAB66031-B66057,
CC   AAB66064-B66083 and AAB66085). The TANGO/INTERCEPT proteins and coding
CC   sequences are useful for the treatment of neurological disorders such as
CC   central nervous system (CNS) disorders, CNS-related disorders, focal
CC   brain disorders, global-diffuse cerebral disorders and other
CC   neurological and cerebrovascular disorders. The CNS disorders include
CC   Alzheimer's disease, senile dementia, Huntington's disease, amyotrophic
CC   lateral sclerosis, Parkinson's, Gilles de la Tourette's syndrome,
CC   autonomic function disorders such as hypertension and sleep disorders,
CC   neuropsychiatric disorders, psychoactive substance use disorders,
CC   anxiety, and bipolar affective disorder.
XX
SQ   Sequence 2044 BP; 618 A; 401 C; 460 G; 565 T; 0 other;


  Query Match             100.0%;  Score 1269;  DB 22;  Length 2044;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1269;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 atgttggaaaccttgtcaagacagtggattgtctcacacagaatggaaatgtggcttctg 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     126 atgttggaaaccttgtcaagacagtggattgtctcacacagaatggaaatgtggcttctg 185

Qy      61 attctggtggcgtatatgttccagagaaatgtgaattcagtacatatgccaactaaagct 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     186 attctggtggcgtatatgttccagagaaatgtgaattcagtacatatgccaactaaagct 245

Qy     121 gtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgt 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     246 gtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgt 305

Qy     181 gaggaatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctcga 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     306 gaggaatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctcga 365

Qy     241 ggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggccta 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     366 ggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggccta 425

Qy     301 gttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctg 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     426 gttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctg 485

Qy     361 gcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaa 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     486 gcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaa 545

Qy     421 cacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatggct 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     546 cacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatggct 605

Qy     481 aggtttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaaagatc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     606 aggtttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaaagatc 665

Qy     541 tattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgcca 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     666 tattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgcca 725

Qy     601 gagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcat 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     726 gagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcat 785

Qy     661 gcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgttt 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     786 gcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgttt 845

Qy     721 ggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgt 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     846 ggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgt 905

Qy     781 ggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaac 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     906 ggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaac 965

Qy     841 accaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaaca 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     966 accaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaaca 1025

Qy     901 tctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcattt 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1026 tctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcattt 1085

Qy     961 gactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtac 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1086 gactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtac 1145

Qy    1021 agagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttca 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1146 agagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttca 1205

Qy    1081 aatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaat 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1206 aatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaat 1265

Qy    1141 attcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtac 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1266 attcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtac 1325

Qy    1201 aatgaaatcatccatctgatgcagcaggaggagaccaacctttcccagggacggtgtgag 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1326 aatgaaatcatccatctgatgcagcaggaggagaccaacctttcccagggacggtgtgag 1385

Qy    1261 gccgtattg 1269
          |||||||||
Db    1386 gccgtattg 1394


RESULT   3
AAQ42310
ID   AAQ42310 standard; DNA; 1378 BP.
XX
AC   AAQ42310;
XX
DT   20-SEP-1993  (first entry)
XX
DE   RGL precursor.
XX
KW   Rabbit gastric lipase; RGL; pRGLN2.1; fat; bioconversion;
KW   hydrolysis; transesterification; ds.
XX
OS   Oryctolagus cuniculus.
XX
FH   Key             Location/Qualifiers
FT   misc_RNA        120..1247
FT                   /*tag= a
FT                   /note= "claim 1; page 10-11"
XX
PN   EP542629-A.
XX
PD   19-MAY-1993.
XX
PF   12-NOV-1992;   92EP-0403055.
XX
```

```
PR   13-NOV-1991;   91FR-0013948.
XX
PA   (LJOU ) INST RECH JOUVEINAL.
XX
PI   Benicourt C,  Blanchard C,  Junien J;
XX
DR   WPI; 1993-161080/20.
DR   P-PSDB; AAR37302.
XX
PT   Rabbit gastric lipase, its precursor and their DNA - useful for
PT   treating conditions linked to gastric lipase deficiency, such as
PT   mucoviscidiosis and pancreatic exocrine insufficiency
XX
PS   Claim 1; Fig 6; 31pp; French.
XX
CC   mRNA was isolated from rabbit gastric mucosal cells, converted to
CC   cDNA and fragments cloned in pUC18.  Recombinant plasmids were used
CC   to transform E.coli MM294 and cells screened for reactivity with
CC   probes corresp. to known parts of the RGL sequence.  One positive
CC   clone contd. plasmid pJ0101 which contd. a 1.35 kb PstI sequence.
CC   This fragment was isolated, subjected to PCR and the amplification
CC   prod. ligated with the 2.2 kb BglII-SacI fragment of pRU276 (contg.
CC   a synthetic Tac promoter and transcription stop signal).  The
CC   recombinant plasmid was designated pRGLN2.1.
XX
SQ   Sequence 1378 BP; 397 A; 287 C; 276 G; 418 T; 0 other;


  Query Match             38.6%;  Score 489.2;  DB 14;  Length 1378;
  Best Local Similarity   65.0%;  Pred. No. 2.7e-145;
  Matches  722;  Conservative    0;  Mismatches  388;  Indels    0;  Gaps    0;

Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
          |||| ||||     |||||||||||| | || ||     |    || || ||  ||||
Db    136 accctgaagtgaatatgaatattagtcagatgatttcctactggggatacccaagtgaaa 195

Qy    185 aatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctcgaggcc 244
          |||||||||| | ||||||||||| |||||||||   |||||||| ||||||   ||
Db    196 aatatgaagttgtgactgaagatggttatatccttgaggttaacagaattccttatggga 255

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
              |  |   ||| | |||    || || || || ||  ||||||||||  |  |||
Db    256 agaaaaattcagggaacagaggccagagacccgtcgtatttttgcagcatggtttgcttg 315

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
            |   ||    |||||||| |||||||||||||||||| ||||||| ||| ||||||||||
Db    316 catcagcctcaaactggatctccaacctgcccaacaacagcctggcctttattctggcag 375

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
          ||||||||| ||  |||||| |||||||||||||| |||||  |||||||  ||| | ||
Db    376 atgctggttatggtgtgtggctggggaacagcagaggaaatacctggtccagaagaaact 435

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
                ||   ||||   | ||| |||||||||| || | |||||| |||||||  |
Db    436 tatactattcaccagactccgttgaattctgggcttttagctttgatgaaatggctaaat 495

Qy    485 ttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaaagatctatt 544
          |||||||| ||    ||  |||| ||| |  || |||| || |||||||| |  | |
Db    496 atgaccttccagccacaattgacttcattgtaaaggaaactggacaggaaaagctgcact 555

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcattttcacccatgccagagc 604
          |||| ||  |||| |||||||||||||| || |||||||| ||||| || |  ||  |||
Db    556 atgttggtcattcccagggcaccaccattggttttattgccttttctacaaatcccaagc 615

Qy    605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
          ||||| | | ||||||||  | || ||| ||||| ||  | |||||||| ||| || |||
Db    616 tggctgaaagaatcaaaaccttttatgcattagctccagtcgccactgtgaagtatacaa 675

Qy    665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
          ||||||  |  | |||| || ||| | ||   |  ||| |||||    | |||||
Db    676 aaagccttgtaaacaaacttaggtttattcctccaaccatgttcaagattatatttggtg 735

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
```

```
          | |||  |||       || |     |||     |   ||||  | | |  | |||  ||
Db    736 acaaaatattctacccacacaatttctttgatcaatttcttgccactcaagtgtgttccc 795

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
            | ||  ||  ||   ||||| || |||  | | ||  |  |  |||| ||  ||| |
Db    796 gtgagacactgaatgtaatttgcagcaatgccttatttatcatttgtggctttgacagcg 855

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
            ||  ||||||||||| ||      ||| || |   | || | ||  || ||||| || |
Db    856 caaacttgaacatgagtcgcttggatgtgtacgtgtcacataatccggcaggaacttcag 915

Qy    905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
          | ||||| || || ||||||| |||||| || || |||||  |  | |  || ||| | |
Db    916 ttcaaaacatgctgcactggacccaggctgttaaatctggaaattttcaagcttttaatt 975

Qy    965 ggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
          ||||  |||     |  | ||| | |    | | |||||||||| || |||      ||||  |
Db    976 ggggaagtccagctcagaatgtagtgcacttcaatcagcccacacctccctactacaatg 1035

Qy    1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
           | |  |  ||||  || || |  ||| ||||||  || ||  | |||||| |  |  | |
Db    1036 tgaccgccatgaatgtgccaattgcagtgtggagtggtggcaatgactggttggctgacc 1095

Qy    1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
           | |||| ||  |  |  ||||  |  |  |  | || ||||| ||||| ||| | ||||
Db    1096 cccaagatgttgacctttgcttccaaaactttctaatctcatttaccacaaggagattc 1155

Qy    1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
             |  ||    |||| ||||||| |||||||   || ||||||||||     | |||||||
Db    1156 ttccatataatcacttggatttttatctgggcaatgaatgctcctcaagaagtttacaatg 1215

Qy    1205 aaatcatccatctgatgcagcaggaggaga 1234
          |||  ||   | |||||     | ||  | |
Db    1216 aaattatttctatgatggcaaaagataaaa 1245


RESULT   4
AAN60685
ID   AAN60685 standard; DNA; 1367 BP.
XX
AC   AAN60685;
XX
DT   08-JUN-1991  (first entry)
XX
DE   Sequence encoding pregastric lipase.
XX
KW   Precursor polypeptide; secretion vector; enzyme; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   sig_peptide     47..103
FT                   /*tag=  a
FT   mat_peptide     104..1243
FT                   /*tag=  b
XX
PN   WO8603778-A.
XX
PD   03-JUL-1986.
XX
PF   23-DEC-1985;   85WO-GB00599.
XX
PR   21-DEC-1984;   84GB-0032483.
PR   23-DEC-1985;   85WO-GB00599.
PR   01-JAN-1986;   86GB-0019568.
XX
PA   (BREW-) BREWING RES FOUND.
PA   (TUBB/) TUBB R S.
XX
PI   Tubb RS;
XX
DR   WPI; 1986-182910/28.
DR   P-PSDB; AAP60724.
```

```
XX
PT   New precursor polypeptide of defined sequence - and corresp. DNA
PT   used to transform hosts for prodn. of the polypeptide
XX
PS   Example; Fig 5; 60pp; English.
XX
CC   The inventors claim a DNA sequence encoding AAP60723 linked to a
CC   promoter upstream and a gene for a polypeptide downstream.
CC   Particular examples are the yeast emzyme AMG, the mammalian enzyme,
CC   gastric lipase and the mammalian lymphokine, interferon-alpha2.
XX
SQ   Sequence 1367 BP; 397 A; 284 C; 269 G; 417 T; 0 other;

  Query Match           37.0%;  Score 470;  DB 7;  Length 1367;
  Best Local Similarity   64.2%;  Pred. No. 3.5e-139;
  Matches  707;  Conservative    0;  Mismatches  395;  Indels    0;  Gaps    0;

Qy    119 ctgtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccct 178
Db    123 ctggaagccctgaagtgactatgaacattagtcagatgattacttattggggatacccaa 182

Qy    179 gtgaggaatatgaagtcgcaactgaagatgggtatatccctttctgttaacaggattcctc 238
Db    183 atgaagaatatgaagttgtgactgaagatggttatattcttgaagtcaatagaattcctt 242

Qy    239 gaggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcc 298
Db    243 atgggaagaaaaattcagggaatacaggccagagacctgttgtgtttttgcagcatggtt 302

Qy    299 tagttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattc 358
Db    303 tgcttgcatcagccacaaactggatttccaacctgccgaacaacagccttgccttcattc 362

Qy    359 tggcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaa 418
Db    363 tggcagatgctggttatgatgtgtggctgggcaacagcagaggaaacacctgggccagaa 422

Qy    419 aacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatgg 478
Db    423 gaaacttgtactattcaccagattcagttgaattctgggctttcagctttgatgaaatgg 482

Qy    479 ctaggtttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaaga 538
Db    483 ctaaatatgaccttccagccacaatcgacttcattgtaaagaaaactggacagaagcagc 542

Qy    539 tctattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgc 598
Db    543 tacactatgttggccattcccagggcaccaccattggttttattgccttttccaccaatc 602

Qy    599 cagagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagc 658
Db    603 ccagcctggctaaaagaatcaaaaccttctatgctctagctcctgttgccactgtgaagt 662

Qy    659 atgcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgt 718
Db    663 atacaaaaagccttataaacaaacttagatttgttcctcaatccctcttcaagtttatat 722

Qy    719 ttggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttt 778
Db    723 ttggtgacaaaatattctacccacacaacttctttgatcaatttcttgctactgaagtgt 782

Qy    779 gtggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattca 838
Db    783 gctcccgtgagatgctgaatctcctttgcagcaatgccttatttataatttgtggatttg 842

Qy    839 acaccaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaa 898
Db    843 acagtaagaactttaacacgagtcgcttggatgtgtatctatcacataatccagcaggaa 902

Qy    899 catctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcat 958
Db    903 cttctgttcaaaacatgttccattggacccaggctgttaagtctgggaaattccaagctt 962

Qy    959 ttgactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggt 1018
Db    963 atgactggggaagcccagttcagaataggatgcactatgatcagtcccaacctccctact 1022

Qy   1019 acagagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggcttt 1078
Db   1023 acaatgtgacagccatgaatgtaccaattgcagtgtggaacggtggcaaggacctgttgg 1082

Qy   1079 caaatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataaga 1138
Db   1083 ctgacccccaagatgttggccttttgcttccaaaactccccaatcttatttaccacaagg 1142

Qy   1139 atattcctgaatgggctcacgtggatttcatctgggggtttggatgctcctcaccgtatgt 1198
Db   1143 agattccttttacaatcacttggactttatctgggcaatggatgcccctcaagaagttt 1202

Qy   1199 acaatgaaatcatccatctgat 1220
Db   1203 acaatgacattgtttctatgat 1224


RESULT   5
AAN60566
ID   AAN60566 standard; DNA; 1367 BP.
XX
AC   AAN60566;
XX
DT   22-AUG-1991  (first entry)
XX
DE   Sequence encoding human pregastric lipase.
XX
KW   Cystic fibrosis therapy; enzyme; lipase deficiency; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             47..103
FT                   /*tag=  a
FT   mat_peptide     104..1243
FT                   /*tag=  b
XX
PN   WO8601532-A.
XX
PD   13-MAR-1986.
XX
PF   15-AUG-1985;   85WO-GB00364.
XX
PR   21-AUG-1984;   84GB-0021210.
PR   15-AUG-1985;   85WO-GB00364.
PR   01-JAN-1986;   86GB-0008897.
XX
PA   (CELL-) CELLTECH LTD.
PA   (LOWE/) LOWE P A.
XX
PI   Lowe PA;
XX
DR   WPI; 1986-081634/12.
DR   P-PSDB; AAP60658.
XX
PT   New gastric lipase protein, esp. of human origin - for treating
PT   lipase deficiency, and DNA sequences coding for it
XX
PS   Disclosure; Fig 3; 39pp; English.
XX
CC   The inventors claim a pregastric lipase protein and a gene encoding
CC   it. Gastric lipase protein is useful for oral administration to
CC   treat lipase deficiency, e.g. cystic fibrosis or pancreatitis.
XX
SQ   Sequence 1367 BP; 397 A; 284 C; 269 G; 417 T; 0 other;

  Query Match           37.0%;  Score 470;  DB 7;  Length 1367;
```

```
  Best Local Similarity   64.2%;  Pred. No. 3.5e-139;
  Matches  707;  Conservative    0;  Mismatches  395;  Indels    0;  Gaps    0;

Qy    119 ctgtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccct 178
          |||    ||| ||||     ||||| ||||||  | || ||     ||   || || ||
Db    123 ctggaagccctgaagtgactatgaacattagtcagatgattacttattggggatacccaa 182

Qy    179 gtgaggaatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctc 238
          ||| ||||||||||| |  ||||||||||| ||||| |||   || || || ||||||
Db    183 atgaagaatatgaagttgtgactgaagatggttatattcttgaagtcaatagaattcctt 242

Qy    239 gaggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcc 298
            ||       |  |   ||| |||||    || ||||| |||||  ||||||||||
Db    243 atgggaagaaaaattcagggaatacaggccagagacctgttgtgtttttgcagcatggtt 302

Qy    299 tagttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattc 358
          |  ||| |   || |  |||||||||||||||||||| ||||| ||||| | ||||||||
Db    303 tgcttgcatcagccacaaactggatttccaacctgccgaacaacagccttgccttcattc 362

Qy    359 tggcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaa 418
          |||||||||||||| ||| |||||| |||| |||||||| |||||| |||| |   |||
Db    363 tggcagatgctggttatgatgtgtggctgggcaacagcagaggaaacacctgggccagaa 422

Qy    419 aacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatgg 478
          | ||  |      ||   |||   || ||| |||||||||||||| | |||||| ||||
Db    423 gaaacttgtactattcaccagattcagttgaattctgggctttcagctttgatgaaatgg 482

Qy    479 ctaggtttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaaaga 538
          |||  | |||||||||| ||    ||  |||| ||| |  ||||||| || ||| |  ||
Db    483 ctaaatatgaccttccagccacaatcgacttcattgtaaagaaaactggacagaagcagc 542

Qy    539 tctattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgc 598
          |  | ||||| ||| |||| |||||||||||||| || |||||||| ||||||||||  |
Db    543 tacactatgttggccattcccagggcaccaccattggttttattgccttttccaccaatc 602

Qy    599 cagagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagc 658
          |    |||||| | | ||||||||  |  | |||| |||| ||  | ||||||||  |||
Db    603 ccagcctggctaaaagaatcaaaaccttctatgctctagctcctgttgccactgtgaagt 662

Qy    659 atgcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgt 718
          || |||||||||     | |||| ||   ||  | ||  |     |  |||||    | |
Db    663 atacaaaaagccttataaacaaacttagatttgttcctcaatccctcttcaagtttatat 722

Qy    719 ttggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttt 778
          ||||  | |||  |||        || | |   |||     |   |||| || | |  | |
Db    723 ttggtgacaaaatattctacccacacaacttctttgatcaatttcttgctactgaagtgt 782

Qy    779 gtggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattca 838
          |   ||  | ||| ||  |||  |||| || |||  | | ||  |  |  |||||||
Db    783 gctcccgtgagatgctgaatctcctttgcagcaatgccttatttataatttgtggatttg 842

Qy    839 acaccaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaa 898
          |||  || ||  | |||| ||| ||      ||| |||    | || | ||  || ||||
Db    843 acagtaagaactttaacacgagtcgcttggatgtgtatctatcacataatccagcaggaa 902

Qy    899 catctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcat 958
          | ||||| ||||| ||  |  || |||| |||||| || || |||||  || |||  || |
Db    903 cttctgttcaaaacatgttccattggacccaggctgttaagtctgggaaattccaagctt 962

Qy    959 ttgactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggt 1018
          |||||||||  ||        | |||  |    | |   |||||  |    |||     |
Db    963 atgactggggaagcccagttcagaataggatgcactatgatcagtcccaacctccctact 1022

Qy   1019 acagagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggcttt 1078
          |||  || | ||  ||||  || || |  ||| ||||||  || ||  |||||  | |
Db   1023 acaatgtgacagccatgaatgtaccaattgcagtgtggaacggtggcaaggacctgttgg 1082

Qy   1079 caaatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataaga 1138
          |  | ||  |||| ||     |  |||| |  |  |  |||| || || ||||| |||
Db   1083 ctgacccccaagatgttggccttttgcttccaaaactccccaatcttatttaccacaagg 1142

Qy   1139 atattcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgt 1198
          | ||||||   |  . |||| |||| || |||||||   |||||||  |||||    | |

Db   1143 agattccttttttacaatcacttggactttatctgggcaatggatgcccctcaagaagttt 1202

Qy   1199 acaatgaaatcatccatctgat 1220
          |||||||  ||  |   | ||||
Db   1203 acaatgacattgtttctatgat 1224


RESULT   6
AAT58916
ID   AAT58916 standard; DNA; 1367 BP.
XX
AC   AAT58916;
XX
DT   19-AUG-1997  (first entry)
XX
DE   Human gastric lipase coding sequence.
XX
KW   Duodenal; gastric; lipase; transgenic; plant; recombinant; extract; ss;
KW   food; absorption; fat; pancreatic; cystic fibrosis; exocrine; dairy;
KW   hydrolysis; trans-esterification; substrate; enzyme; biofuel; human.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             47..1243
FT                   /*tag= a
FT                   /product= gastric lipase
FT   sig_peptide     47..103
FT                   /*tag= b
FT   mat_peptide     104..1240
FT                   /*tag= c
XX
PN   WO9633277-A2.
XX
PD   24-OCT-1996.
XX
PF   19-APR-1996;   96WO-FR00606.
XX
PR   20-APR-1995;   95FR-0004754.
XX
PA   (BIOC-) BIOCEM SA.
PA   (LJOU ) INST RECH JOUVEINAL.
XX
PI   Baudino S,  Benicourt C,  Cudrey C,  Gruber V,  Lenee P;
PI   Merot B;
XX
DR   WPI; 1996-485783/48.
DR   P-PSDB; AAW09383.
XX
PT   Recombinant human or canine pre-duodenal lipase prodn. in transgenic
PT   plants - useful for facilitating absorption of fat, as bio-catalysts
PT   and for prodn. of bio-fuel
XX
PS   Claim 3; Fig 4; 130pp; French.
XX
CC   This is the nucleotide sequence encoding the human pre-duodenal (i.e.
CC   gastric) lipase enzyme.  The sequence can be used to generate transgenic
CC   plants producing recombinant lipase in an enzymatically active form.
CC   Alternatively bases encoding amino acids 20-23 or 20-73 (1-4 or 1-54 of
CC   the mature protein, respectively) can be deleted to form the derivatives
CC   designated delta-4 or delta-54 respectively.  Plants, or their extracts,
CC   expressing the lipases or the truncated derivatives, can be used:
CC   (a) as pharmaceuticals or food to facilitate absorption of fat, either
CC   in healthy subjects or in patients with inadequate levels of
CC   gastric/pancreatic lipase e.g. those with cystic fibrosis or exocrine
CC   pancreatic insufficient, in the elderly or in patients undergoing medical
CC   treatment which alters fat adsorption; (b) for performing industrial or
CC   agricultural reactions, e.g. in processing of fats or in the dairy
CC   industry, for hydrolysis or transesterification reactions, etc., where
CC   the plant material may provide both enzyme and substrate.  The transgenic
CC   plants can also be used for biofuel production.
XX
SQ   Sequence 1367 BP; 397 A; 284 C; 269 G; 417 T; 0 other;
```

```
  Query Match          37.0%;  Score 470;  DB 17;  Length 1367;
  Best Local Similarity  64.2%;  Pred. No. 3.5e-139;
  Matches  707;  Conservative    0;  Mismatches  395;  Indels    0;  Gaps    0;

Qy     119 ctgtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccct 178
           |||    ||| ||||     ||||| ||||||  | || ||     ||   || || ||
Db     123 ctggaagccctgaagtgactatgaacattagtcagatgattacttattggggatacccaa 182

Qy     179 gtgaggaatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctc 238
           ||| ||||||||||| |  |||||||||| ||||| |||   || || || ||||||
Db     183 atgaagaatatgaagttgtgactgaagatggttatattcttgaagtcaatagaattcctt 242

Qy     239 gaggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcc 298
             ||       |  |   ||| |||||     || ||||| |||||  ||||||||||
Db     243 atgggaagaaaaattcagggaatacaggccagagacctgttgtgtttttgcagcatggtt 302

Qy     299 tagttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattc 358
           |  ||| |   || |  ||||||||||||||||||||.||||| ||||| | ||||||||
Db     303 tgcttgcatcagccacaaactggatttccaacctgccgaacaacagccttgccttcattc 362

Qy     359 tggcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaa 418
           ||||||||||||||| ||| |||||| |||| |||||||| |||||| |||| |  |||
Db     363 tggcagatgctggttatgatgtgtggctgggcaacagcagaggaaacacctgggccagaa 422

Qy     419 aacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatgg 478
           | ||  |      ||   |||   || ||| ||||||||||||| | |||||| ||||
Db     423 gaaacttgtactattcaccagattcagttgaattctgggctttcagctttgatgaaatgg 482

Qy     479 ctaggtttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaaaga 538
           |||  | |||||||||| ||    ||  |||| ||| |  |||||| || ||| |  ||
Db     483 ctaaatatgaccttccagccacaatcgacttcattgtaaagaaaactggacagaagcagc 542

Qy     539 tctattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgc 598
           |  | ||||| ||| |||| |||||||||||||| || |||||||| ||||||||||  |
Db     543 tacactatgttggccattcccagggcaccaccattggttttattgccttttccaccaatc 602

Qy     599 cagagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagc 658
           |    |||||| | | ||||||||  |  | |||| |||| ||  | ||||||||| |||
Db     603 ccagcctggctaaaagaatcaaaaccttctatgctctagctcctgttgccactgtgaagt 662

Qy     659 atgcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgt 718
           || |||||||||     | |||| ||   ||  | ||  |     |  |||||    | |
Db     663 atacaaaaagccttataaacaaacttagatttgttcctcaatccctcttcaagtttatat 722

Qy     719 ttggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttt 778
           ||||  | |||  |||       || | |   |||     |   |||| || | |  | |
Db     723 ttggtgacaaaatattctacccacacaacttctttgatcaatttcttgctactgaagtgt 782

Qy     779 gtggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattca 838
           |   ||  | ||| ||  |||   |||| || |||  | | ||  |  |  |||||||
Db     783 gctcccgtgagatgctgaatctcctttgcagcaatgccttatttataatttgtggatttg 842

Qy     839 acaccaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaa 898
           |||  || ||  | |||| ||| ||       ||| |||    | || | ||  || ||||
Db     843 acagtaagaactttaacacgagtcgcttggatgtgtatctatcacataatccagcaggaa 902

Qy     899 catctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcat 958
           | ||||| ||||| ||  | || |||| |||||| || || ||||  || |||  || |
Db     903 cttctgttcaaaacatgttccattggacccaggctgttaagtctgggaaattccaagctt 962

Qy     959 ttgactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggt 1018
           |||||||||| ||        | |||  |    | |  |||||  |    |||      |
Db     963 atgactggggaagcccagttcagaataggatgcactatgatcagtcccaacctccctact 1022

Qy    1019 acagagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggcttt 1078
           |||  || | ||  ||||  || || |  ||| ||||||  || ||  |||||  | |
Db    1023 acaatgtgacagccatgaatgtaccaattgcagtgtggaacggtggcaaggacctgttgg 1082

Qy    1079 caaatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataaga 1138
           |  | ||  |||| ||      |  ||||  |  |  |  |||| || || ||||| |||
Db    1083 ctgacccccaagatgttggccttttgcttccaaaactccccaatcttatttaccacaagg 1142
```

```
Qy    1139 atattcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgt 1198
           | ||||||   |    |||| |||| || |||||||   |||||| |||||     | |
Db    1143 agattccttttacaatcacttggactttatctgggcaatggatgcccctcaagaagttt 1202

Qy    1199 acaatgaaatcatccatctgat 1220
           ||||||| ||  |   | ||||
Db    1203 acaatgacattgtttctatgat 1224


RESULT   7
AAN50385
ID   AAN50385 standard; DNA; 1336 BP.
XX
AC   AAN50385;
XX
DT   17-JAN-1992  (first entry)
XX
DE   Rat prelingual lipase gene.
XX
KW   Prelingual lipase; enzyme; EC-3.1.1.3; ss.
XX
OS   Rattus rattus.
XX
FH   Key             Location/Qualifiers
FT   CDS             7..1194
FT                   /*tag= a
XX
PN   GB2142337-A.
XX
PD   16-JAN-1985.
XX
PF   29-JUN-1984;   84GB-0016581.
XX
PR   01-JUL-1983;   83GB-0017989.
PR   05-SEP-1983;   83GB-0023759.
XX
PA   (CELL-) CELLTECH LTD.
XX
PI   Carey NH,  Williamson R;
XX
DR   WPI; 1985-014450/03.
DR   P-PSDB; AAP50322.
XX
PT   New lingual lipase protein for treatment of lipase deficiency - also
PT   new pre-lingual lipase protein and related products
XX
PS   Disclosure; Fig 5; 15pp; English.
XX
CC   This sequence may be expressed in a transformant host organism using
CC   a vector plasmid in order to produce a prelingual lipase protein.
CC   The expressed protein may be used for the treatment of lipase
CC   deficiency.
XX
SQ   Sequence 1336 BP; 409 A; 285 C; 263 G; 379 T; 0 other;

  Query Match          36.6%;  Score 464.4;  DB 6;  Length 1336;
  Best Local Similarity  63.4%;  Pred. No. 2.1e-137;
  Matches  711;  Conservative    0;  Mismatches  411;  Indels    0;  Gaps    0;

Qy     119 ctgtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccct 178
           |||   |||| ||||||   |||||||||||| | || ||     |    || ||||| |
Db      80 ctggaaaccctgaagcaaatatgaatattagtcagatgataacttactggggatatccat 139

Qy     179 gtgaggaatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctc 238
           || | |||||||||||  |  |||||||||||| || || ||    ||  |||| |||||||
Db     140 gtcaagaatatgaagttgttactgaagatggctacattctgggggtctacagaattcctc 199

Qy     239 gaggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcc 298
             ||       |  || | || | |||    || ||||||||||   ||||||||||
Db     200 atgggaagaataattctgaaaatataggcaagagacctgtggtgtatttgcagcatggtt 259
```

```
Qy    299 tagttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattc 358
          |  | | |   || |  |||||||||| | || || || ||||| ||||| | ||| || |
Db    260 tgatcgcatcagccacaaactggattgcaaatctaccaaacaacagcctcgcctttatgc 319

Qy    359 tggcagatgctggtttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaa 418
          | ||||| ||||| | ||| |||||| |||||||||| | ||||| | ||||| | |
Db    320 tagcagacgctggctatgatgtgtggctggggaacagtcgaggaaatacatggtccagga 379

Qy    419 aacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatgg 478
          || |         |||   ||||  || ||| ||||||||||||||| | |||||| ||||
Db    380 aaaatgtatactactcaccagactcagttgaattctgggctttcagctttgatgaaatgg 439

Qy    479 ctaggtttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaaaga 538
          |||  | |||||||||| ||    |||||||| ||| | |||||||| || || || ||||
Db    440 ctaaatatgaccttcccgccacaataaacttcattgtacagaaaactggacaagagaaga 499

Qy    539 tctattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgc 598
          |  | ||||| ||  |||| ||| ||||||| || || || ||||| |||| ||||  |
Db    500 tacactatgttggtcattctcagcgcaccactattggtttcattgccttttctaccaatc 559

Qy    599 cagagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagc 658
          |    ||||| | |||||||| | || || ||| |||| ||  | || || || |||
Db    560 ctacactggccaaaaaaatcaagacgttttatgcattagctccagttgctaccgtgaagt 619

Qy    659 atgcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgt 718
          || || |||| |||  ||  ||  |||  ||  | ||     |  |  |||||    |||
Db    620 atacacaaagtcccttgaaaaagatttcatttattcctacatttcttttcaagcttatgt 679

Qy    719 ttggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccttt 778
          ||||||| |||   || |||   || |||   |||          |||| ||   |  | |
Db    680 ttggcaagaaaatgttcctgccccacacctactttgatgactttcttggtaccgaagtgt 739

Qy    779 gtggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattca 838
          |    | || | |||| ||||   | || || || |  | ||  |  |  |||||||
Db    740 gctcacgggaggttctagatcttctctgcagcaacactttattcatcttctgtggatttg 799

Qy    839 acaccaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaa 898
          |||  || ||  | ||  |||| ||      ||| |||  |  || | ||  || || |
Db    800 acaagaaaaacttaaatgtgagtcgttttgatgtgtatctagggcataatccagcaggga 859

Qy    899 catctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcat 958
          ||||||| ||| |  |||| |||||   |||  || |  |||| || | |  || ||  |
Db    860 catctgttcaagactttctccactgggcacagcttgttagatctgggaaatttcaagcct 919

Qy    959 ttgactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggt 1018
          |  ||||||| ||     || | ||  ||  | | | ||| |||  ||| |||     |
Db    920 tcaactggggaagcccatcccagaacatgttacactacaaccagaaaacgcctcctgaat 979

Qy   1019 acagagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggcttt 1078
          |    ||  ||  ||||| || ||    ||| |||||  || ||  | |||   ||
Db    980 atgatgtgtcagccatgactgtgccagttgcagtgtggaacggtggcaatgacatcctgg 1039

Qy   1079 caaatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataaga 1138
          |  ||||  |||| ||    ||||||||  |  |  |  |||||||| | | |||||||
Db   1040 ctgatccccaagatgtcgccatgctgcttcccaaactctccaacctcctgttccataagg 1099

Qy   1139 atattcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgt 1198
          | |||| ||  |    |||| |||| ||||||||||   ||||||| |||||     | |
Db   1100 agattcttgcctacaatcacctggacttcatctgggcaatggatgcccctcaagaggttt 1159

Qy   1199 acaatgaaatcatccatctgatgcagcaggaggagaccaacc 1240
          ||||||| || ||     |||||    | ||  | |  || |
Db   1160 acaatgagatgatttccatgatggcagaagactaaaagaatc 1201


RESULT   8
AAQ68388
ID   AAQ68388 standard; DNA; 1528 BP.
XX
AC   AAQ68388;
XX
DT   20-FEB-1995  (first entry)
XX
DE   Canine gastric lipase.
XX
KW   Canine gastric lipase; CGL; dog; fat; mucoviscidosis; enzyme;
KW   bioconversion; exocrine pancreatic insufficiency; ss.
XX
OS   Canis familiaris.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1140
FT                   /*tag= a
FT   misc_RNA        1..1137
FT                   /*tag= b
FT                   /note= "claim 1"
XX
PN   WO9413816-A.
XX
PD   23-JUN-1994.
XX
PF   16-DEC-1993;   93WO-FR01260.
XX
PR   16-DEC-1992;   92FR-0015201.
XX
PA   (LJOU ) INST RECH JOUVEINAL.
XX
PI   Benicourt C,  Blanchard C,  Junien J;
XX
DR   WPI; 1994-217890/26.
DR   P-PSDB; AAR56870.
XX
PT   Recombinant canine gastric lipase and nucleic acid encoding it -
PT   are used for improving absorption of ingested fat, treating
PT   mucoviscidosis etc. and in enzymatic bio-conversions
XX
PS   Claim 1; Fig 8; 52pp; French.
XX
CC   The sequence given below is the sequence of figure 8, altered
CC   according to the amendments described on page 2 of the appended
CC   letter.
CC   CGL is used to improve absorption of ingested fat, in healthy and
CC   sick patients (e.g. having altered levels of gastric lipase); to
CC   treat conditions associated with insufficiency (or lack) of lipases,
CC   esp. mucoviscidosis or exocrine pancreatic insufficiency and partic.
CC   where immobilised, for bioconversions, e.g. hydrolysis or
CC   transesterification (other mammalian gastric lipases, or derivs.,
CC   can be used in this application).
XX
SQ   Sequence 1528 BP; 394 A; 386 C; 328 G; 420 T; 0 other;


  Query Match             35.0%;  Score 444.2;  DB 15;  Length 1528;
  Best Local Similarity   62.8%;  Pred. No. 6.1e-131;
  Matches  689;  Conservative    0;  Mismatches  408;  Indels    0;  Gaps    0;

Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
          |||| ||||    ||||||||| ||| | || |||    |    || || ||   |||||
Db     26 acccgaagtgaccatgaatataagtcagatgatcacctactggggatacccagctgagg 85

Qy    185 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 244
          |||||||||| |  || ||||| || |||||||||    |  |||| ||||||   ||
Db     86 aatatgaagttgtgaccgaagacggttatatccttgggatcgacagaattccttatggga 145

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
             |  |  |||| | |||    || ||||| |  ||  |||| || ||  |  | |
Db    146 ggaaaaattcagagaatataggccggagacctgttgcatttttgcaacacggtttgctcg 205

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
          |   || |  |||||||| |||||||||||||||||| ||||||| |||||| |||| |
Db    206 catcagccacaaactggatctccaacctgcccaacaacagcctggccttcatcctggccg 265

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
          | || || |  |||||||||| |||||||||||||||||| ||| ||||| |  | |   |
Db    266 acgccgggtacgacgtgtggctggggaacagcaggggcaacacctgggccaggaggaatc 325
```

```
Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
          |     |||    |||    |  || ||||||||||||||| | ||| ||||||||||  |
Db    326 tgtactactcgcccgactccgtcgaattctgggctttcagctttgacgagatggctaaat 385

Qy    485 ttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 544
          ||||||||| ||    ||  |||| || ||| |||||||||| ||||| ||| |  | ||
Db    386 atgaccttcccgccaccattgacttcatcttgaagaaaacgggacaggacaagctacact 445

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
          | || ||| |||| |||||||||||||| || || || || ||||||||||  ||  |||
Db    446 acgttggccattcccagggcaccaccattggtttcatcgccttttccaccaatcccaagc 505

Qy    605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
          ||||  |    |||||||  |  | ||| |||||| ||| | ||||| || ||| |  |
Db    506 tggcgaaacggatcaaaaccttctatgcattagctcccgttgccaccgtgaagtacaccg 565

Qy    665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
          ||| ||     | |||| |  || |  | ||     |  |  |||||    | ||||| |
Db    566 aaaccctgttaaacaaactcatgctcgtcccttcgttcctcttcaagcttatatttggaa 625

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
          | |||  |||       ||   |   |||     |   || |  |   |  | ||   ||
Db    626 acaaaatattctacccacaccacttctttgatcaatttctcgccaccgaggtatgctccc 685

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
            | ||   | ||||   | || || ||   | ||||  |  |  |||||||  |||| |
Db    686 gcgagacggtggatctcctctgcagcaacgccctgtttatcatttgtggatttgacacta 745

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
            ||  ||||||||| ||      ||| |||    | || | ||  || |||||||| |
Db    746 tgaacttgaacatgagtcgcttggatgtgtatctgtcacataatccagcaggaacatcgg 805

Qy    905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
          | || ||  | || ||||||  ||||||| || || |||||  |  |||  || |||||||
Db    806 ttcagaacgtgctccactggtcccaggctgttaagtctgggaagttccaagcttttgact 865

Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
          |||| ||        | ||  ||    | |   |||||   |  |||      ||||
Db    866 ggggaagcccagttcagaacatgatgcactatcatcagagcatgcctccctactacaacc 925

Qy   1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
          | | ||| |||   || || |  ||| ||||||  || ||  | |||| |||  |  | |
Db    926 tgacagacatgcatgtgccaatcgcagtgtggaacggtggcaacgacttgctggccgacc 985

Qy   1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
          |  | || ||  |  |  |||| ||  || |  |||| ||||| ||||| | ||| ||||
Db    986 ctcacgatgttgacctttgctttccaagctccccaatctcatttaccacaggaagattc 1045

Qy   1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
          ||   |    |||| |||| || |||||||   |||||||  |||||     | |||||||
Db   1046 ctccttacaatcacttggactttatctgggccatggatgcccctcaagcggtttacaatg 1105

Qy   1205 aaatcatccatctgatg 1221
          ||||  |     |||||
Db   1106 aaattgtttccatgatg 1122
```

```
RESULT   9
AAT58915
ID   AAT58915 standard; cDNA to mRNA; 1528 BP.
XX
AC   AAT58915;
XX
DT   19-AUG-1997  (first entry)
XX
DE   Dog gastric lipase coding sequence.
XX
KW   Dog; duodenal; gastric; lipase; transgenic; plant; recombinant; extract;
KW   food; absorption; fat; pancreatic; cystic fibrosis; exocrine; dairy;
KW   hydrolysis; trans-esterification; substrate; enzyme; biofuel; ss.
XX
OS   Canis familiaris.
XX
```

```
FH   Key             Location/Qualifiers
FT   CDS             1..1140
FT                   /*tag= a
FT                   /product= gastric lipase
XX
PN   WO9633277-A2.
XX
PD   24-OCT-1996.
XX
PF   19-APR-1996;   96WO-FR00606.
XX
PR   20-APR-1995;   95FR-0004754.
XX
PA   (BIOC-) BIOCEM SA.
PA   (LJOU ) INST RECH JOUVEINAL.
XX
PI   Baudino S,  Benicourt C,  Cudrey C,  Gruber V,  Lenee P;
PI   Merot B;
XX
DR   WPI; 1996-485783/48.
DR   P-PSDB; AAW09382.
XX
PT   Recombinant human or canine pre-duodenal lipase prodn. in transgenic
PT   plants - useful for facilitating absorption of fat, as bio-catalysts
PT   and for prodn. of bio-fuel
XX
PS   Claim 2; Fig 1; 130pp; French.
XX
CC   This is the nucleotide sequence encoding the dog pre-duodenal (i.e.
CC   gastric) lipase enzyme.  The sequence can be used to generate transgenic
CC   plants producing recombinant lipase in an enzymatically active form.
CC   Alternatively the first 12 bases of this sequence (encoding amino acids
CC   1-4 of the protein encoded by this sequence) or bases 1-162 of this
CC   sequence (encoding amino acids 1-54) can be deleted to form the
CC   derivatives designated delta-4 or delta-54 respectively.  Plants, or
CC   their extracts, expressing the lipases or the truncated derivatives, can
CC   be used: (a) as pharmaceuticals or food to facilitate absorption of fat,
CC   either in healthy subjects or in patients with inadequate levels of
CC   gastric/pancreatic lipase e.g. those with cystic fibrosis or exocrine
CC   pancreatic insufficient, in the elderly or in patients undergoing medical
CC   treatment which alters fat adsorption; (b) for performing industrial or
CC   agricultural reactions, e.g. in processing of fats or in the dairy
CC   industry, for hydrolysis or transesterification reactions, etc., where
CC   the plant material may provide both enzyme and substrate.  The transgenic
CC   plants can also be used for biofuel production.
XX
SQ   Sequence 1528 BP; 392 A; 387 C; 329 G; 420 T; 0 other;
```

```
  Query Match             35.0%;  Score 444.2;  DB 17;  Length 1528;
  Best Local Similarity   62.8%;  Pred. No. 6.1e-131;
  Matches  689;  Conservative    0;  Mismatches  408;  Indels    0;  Gaps    0;

Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
          |||| ||||    |||||||||.||| | || |||    |    || || ||   |||||
Db     26 accctgaagtgaccatgaatataagtcagatgatcacctactggggatacccagctgagg 85

Qy    185 aatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctcgaggcc 244
          |||'|||||| |' || ||||| || |||||||||    |  |||| |.|||||   ||
Db     86 aatatgaagttgtgaccgaagacggttatatccttgggatcgacagaattccttatggga 145

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
              |  |  |||| | |||    || ||||| |  ||  |||| || ||  |  | |
Db    146 ggaaaaattcagagaatataggccggagacctgttgcatttttgcaacacggtttgctcg 205

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
            |   || |  |||||||| |||||||||||||||||| |||||| |||||| ||||| |
Db    206 catcagccacaaactggatctccaacctgcccaacaaaaagcctggccttcatcctggccg 265

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
          | || || |  |||||||||| |||||||||||||||| ||| ||||| |  | |   |
Db    266 acgccgggtacgacgtgtggctggggaacagcaggggcaacacctgggccaggaggaatc 325
```

```
Qy     425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
           |     |||    |||   |  || ||||||||||||||| | ||| ||||||||||  |
Db     326 tgtactactcgcccgactccgtcgaattctgggctttcagctttgacgagatggctaaat 385

Qy     485 ttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 544
           ||||||||| ||    ||  |||| || ||| |||||||||| ||||| ||| |  | |
Db     386 atgaccttcccgccaccattgacttcatcttgaagaaaacgggacaggacaagctacact 445

Qy     545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
           | || ||| |||| |||||||||||||| || || || || |||||||||. ||  |||
Db     446 acgttggccattcccagggcaccaccattggtttcatcgccttttccaccaatcccaagc 505

Qy     605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
           ||||  |    |||||||  |  | ||| ||||| ||| | ||||| || ||| |  |
Db     506 tggcgaaacggatcaaaaccttctatgcattagctcccgttgccaccgtgaagtacaccg 565

Qy     665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
           ||| ||     | |||| |  || |  | ||     |  |  |||||    | |||||.|
Db     566 aaaccctgttaaacaaactcatgctcgtcccttcgttcctcttcaagcttatatttggaa 625

Qy     725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
           | |||  |||        ||   |  |||     |   || | |   |  | ||  ||
Db     626 acaaaatattctacccacaccacttctttgatcaatttctcgccaccgaggtatgctccc 685

Qy     785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
           | ||   | ||||   | || || ||   | ||||  |  |  |||||||  |||| |
Db     686 gcgagacggtggatctcctctgcagcaacgccctgtttatcatttgtggatttgacacta 745

Qy     845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
           ||  ||||||||||| ||      ||| |||    | || | ||  || |||||||| |
Db     746 tgaacttgaacatgagtcgcttggatgtgtatctgtcacataatccagcaggaacatcgg 805

Qy     905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
           | || ||  | || ||||||  ||||||| || || |||||  |  |||  || |||||||
Db     806 ttcagaacgtgctccactggtcccaggctgttaagtctgggaagttccaagcttttgact 865

Qy     965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
           |||| ||        | ||  ||    | |   |||||   |  |||     ||||
Db     866 ggggaagcccagttcagaacatgatgcactatcatcagagcatgcctccctactacaacc 925

Qy     1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
            | | ||| |||   || || |  ||| ||||||  || ||  | |||| |||  |  | |
Db      926 tgacagacatgcatgtgccaatcgcagtgtggaacggtggcaacgacttgctggccgacc 985

Qy     1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
            |  | || ||  |  |  |||| ||  || |  |||| |||||  ||||| | ||| ||||
Db      986 ctcacgatgttgacctttgctttccaagctccccaatctcatttaccacaggaagattc 1045

Qy     1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
            ||   |    |||| |||| || |||||||   |||||||  |||||    | |||||||
Db     1046 ctccttacaatcacttggactttatctgggccatggatgcccctcaagcggtttacaatg 1105

Qy     1205 aaatcatccatctgatg 1221
            ||||. |     |||||
Db     1106 aaattgtttccatgatg 1122


RESULT 10
AAQ68389
ID   AAQ68389 standard; DNA; 1531 BP.
XX
AC   AAQ68389;
XX
DT   20-FEB-1995  (first entry)
XX
DE   Canine gastric lipase (Met-CGL).
XX
KW   Canine gastric lipase; CGL; dog; fat; mucoviscidosis; enzyme;
KW   bioconversion; exocrine pancreatic insufficiency; ss.
XX
OS   Canis familiaris.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1143
FT                   /*tag= a
XX
PN   WO9413816-A.
XX
PD   23-JUN-1994.
XX
PF   16-DEC-1993;   93WO-FR01260.
XX
PR   16-DEC-1992;   92FR-0015201.
XX
PA   (LJOU ) INST RECH JOUVEINAL.
XX
PI   Benicourt C,  Blanchard C,  Junien J;
XX
DR   WPI; 1994-217890/26.
DR   P-PSDB; AAR56871.
XX
PT   Recombinant canine gastric lipase and nucleic acid encoding it -
PT   are used for improving absorption of ingested fat, treating
PT   mucoviscidosis etc. and in enzymatic bio-conversions
XX
PS   Claim 2; Fig 8; 52pp; French.
XX
CC   The sequence given below is the sequence of figure 8, altered
CC   according to the amendments described on page 2 of the appended
CC   letter.
CC   CGL is used to improve absorption of ingested fat, in healthy and
CC   sick patients (e.g. having altered levels of gastric lipase); to
CC   treat conditions associated with insufficiency (or lack) of lipases,
CC   esp. mucoviscidosis or exocrine pancreatic insufficiency and partic.
CC   where immobilised, for bioconversions, e.g. hydrolysis or
CC   transesterification (other mammalian gastric lipases, or derivs.,
CC   can be used in this application).
XX
SQ   Sequence 1531 BP; 395 A; 386 C; 329 G; 421 T; 0 other;


  Query Match             35.0%;  Score 444.2;  DB 15;  Length 1531;
  Best Local Similarity   62.8%;  Pred. No. 6.1e-131;
  Matches  689;  Conservative    0;  Mismatches  408;  Indels    0;  Gaps    0;

Qy     125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
           |||| ||||    |||||||||| ||| | || |||   |    || || ||   |||||
Db      29 accctgaagtgaccatgaatataagtcagatgatcacctactggggatacccagctgagg 88

Qy     185 aatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctcgaggcc 244
           |||||||||| |  || ||||| || |||||||||    |  |||| ||||||  ||
Db      89 aatatgaagttgtgaccgaagacggttatatccttgggatcgacagaattccttatggga 148

Qy     245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
              |  |||| | |||    || ||||| |  ||  |||| || ||  |  | |
Db     149 ggaaaaattcagagaatataggccggagacctgttgcatttttgcaacacggtttgctcg 208

Qy     305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
           |   || |  |||||||| |||||||||||||||||| ||||||| |||||| ||||| |
Db     209 catcagccacaaactggatctccaacctgcccaacaacagcctggccttcatcctggccg 268

Qy     365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
           | || || |  |||||||| |||||||||||||||| ||| ||||| |  | |   |
Db     269 acgccgggtacgacgtgtggctggggaacagcaggggcaacacctgggccaggaggaatc 328

Qy     425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
           |     |||    |  || |||||||||||||| | ||| ||||||||||  |
Db     329 tgtactactcgcccgactccgtcgaattctgggctttcagctttgacgagatggctaaat 388

Qy     485 ttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 544
           |||||||||| ||    ||  |||| || ||| |||||||||| ||||| ||| |  | |
Db     389 atgaccttcccgccaccattgacttcatcttgaagaaaacgggacaggacaagctacact 448

Qy     545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
           | || ||| |||| |||||||||||||| || || || || ||||||||||  ||  |||
Db     449 acgttggccattcccagggcaccaccattggtttcatcgccttttccaccaatcccaagc 508
```

```
Qy     605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
           ||||  |    |||||||  |  | ||| ||||| ||| | ||||| || |||  |  |
Db     509 tggcgaaacggatcaaaaccttctatgcattagctcccgttgccaccgtgaagtacaccg 568

Qy     665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
           ||| ||     | |||| |  || |  | ||      |  |  |||||    | ||||| |
Db     569 aaaccctgttaaacaaactcatgctcgtcccttcgttcctcttcaagcttatatttggaa 628

Qy     725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
           | |||  |||       ||   |   |||     |   || |  |   |  | ||   ||
Db     629 acaaaatattctacccacaccacttctttgatcaatttctcgccaccgaggtatgctccc 688

Qy     785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
             | ||   | ||||   | || || ||   | ||||  |  |  |||||||  |||| |
Db     689 gcgagacggtggatctcctctgcagcaacgccctgtttatcatttgtggatttgacacta 748

Qy     845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
             ||  |||||||||| ||      ||| |||    | || | ||  || ||||||| |
Db     749 tgaacttgaacatgagtcgcttggatgtgtatctgtcacataatccagcaggaacatcgg 808

Qy     905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
           | || ||  | || ||||||  |||||| || || |||||  |  |||  || |||||||
Db     809 ttcagaacgtgctccactggtcccaggctgttaagtctgggaagttccaagcttttgact 868

Qy     965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
           ||||  ||        | ||  ||    | |   |||||    |  |||      ||||
Db     869 ggggaagcccagttcagaacatgatgcactatcatcagagcatgcctccctactacaacc 928

Qy    1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
           | | ||| |||   || || |  ||| ||||||  || ||  | |||| |||  |  | |
Db     929 tgacagacatgcatgtgccaatcgcagtgtggaacggtggcaacgacttgctggccgacc 988

Qy    1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
           |  | || ||  |  |  |||| ||  || |  |||| ||||| ||||| | ||| ||||
Db     989 ctcacgatgttgacctttgctttccaagctccccaatctcatttaccacaggaagattc 1048

Qy    1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
           ||   |    |||| |||| || ||||||||   ||||||| |||||     | |||||||
Db     1049 ctccttacaatcacttggactttatctgggccatggatgcccctcaagcggtttacaatg 1108

Qy    1205 aaatcatccatctgatg 1221
           ||||  |     |||||
Db     1109 aaattgtttccatgatg 1125
```

RESULT 11
AAF28679
ID AAF28679 standard; cDNA; 699 BP.
XX
AC AAF28679;
XX
DT 05-APR-2001 (first entry)
XX
DE Human protein HP03372 coding sequence #1.
XX
KW Human; hydrophobic domain; immune deficiency; autoimmune disorder;
KW allergy; tissue growth; regeneration; wound healing; burn; tumour;
KW periodontal disease; thrombolytic condition; haemostatic condition;
KW infection; ss.
XX
OS Homo sapiens.
XX
PN WO200102563-A2.
XX
PD 11-JAN-2001.
XX
PF 16-JUN-2000; 2000WO-JP03943.
XX
PR 02-JUL-1999; 99JP-0188835.
XX
PA (SAGA) SAGAMI CHEM RES CENT.
PA (PROT-) PROTEGENE INC.
XX
PI Kato S, Kimura T;
XX
DR WPI; 2001-071581/08.
DR P-PSDB; AAB61608.
XX
PT New human proteins with hydrophobic domains, useful for the treatment
PT of immune disorders, tumors, allergic conditions, thrombosis and
PT microbial infection -
XX
PS Claim 3; Page 113; 153pp; English.
XX
CC The present invention relates to human proteins (AAB61608-AAB61617) and
CC their coding sequences (AAF28679-AAF28698). The proteins of the present
CC invention have hydrophobic domains and can be used for the treatment of
CC various immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), multiple sclerosis, rheumatoid arthritis,
CC autoimmune pulmonary inflammation, graft-versus-host disease and
CC Guillain-Barre syndrome. The proteins may also be useful in the treatment
CC of allergic reactions and conditions, such as asthma and in regulation of
CC haematopoiesis or lymphoid cell deficiencies. The proteins may also have
CC utility in compositions used for bone, cartilage, tendon and/or nerve
CC tissue growth or regeneration as well as wound healing and in the
CC treatment of burns. The proteins may be used in the treatment of
CC periodontal disease and in other tooth repair processes. Other uses
CC include treatment of thrombolytic and haemostatic conditions, treatment
CC or prevention of tumours and inhibiting infection by bacteria, viruses,
CC fungi and other parasites.
XX
SQ Sequence 699 BP; 216 A; 144 C; 148 G; 191 T; 0 other;

Query Match 22.5%; Score 286; DB 22; Length 699;
Best Local Similarity 67.4%; Pred. No. 8.9e-81;
Matches 403; Conservative 0; Mismatches 195; Indels 0; Gaps 0;

```
Qy      86 gaaatgtgaattcagtacatatgccaactaaagctgtggacccagaagcattcatgaata 145
           ||  | || ||       | | |  |   ||   ||   |||| |||||    |||||||
Db      44 gatctatgtatggttatgacaagaaaggaaacaatgcaaaccctgaagctaatatgaata 103

Qy     146 ttactgaaatcatccaacatcaaggctatccctgtgaggaatatgaagtcgcaactgaag 205
           ||||  | || ||      |    || |||| | ||| || ||||| ||  ||||  |||
Db     104 ttagccagattatttcttactggggttatccttatgaagagtatgatgttacaacaaaag 163

Qy     206 atgggtatatcctttctgttaacaggattcctcgaggcctagtgcaacctaagaagacag 265
           |||| |||||||||    || | ||||||| |  ||   ||     ||   || |||||
Db     164 atggttatatccttggaatttataggattccacatggaagaggatgcccagggaggacag 223

Qy     266 gttccaggcctgtggtgttactgcagcatggcctagttggaggtgctagcaactggattt 325
           | | | |||||  ||||   |||||||||||| || ||| | .||| || ||||||||||
Db     224 ctccaaagcctgctgtgtatttgcagcatggcttaattgcatctgccagtaactggattt 283

Qy     326 ccaacctgcccaacaatagcctgggcttcattctggcagatgctggttttgacgtgtgga 385
           ||||||||||||||| ||  |||  ||| |||||||||||  ||||| ||||||||||
Db     284 gcaacctgcccaacaacagtttggctttccttctggcagatagtggttatgacgtgtggt 343

Qy     386 tggggaacagcaggggaaacgcctggtctcgaaaacacaagacactctccatagaccaag 445
           |||||||||| | |||||| | |||||  ||||||||   | | | ||     |   |
Db     344 tggggaacagccgaggaaacacttggtccagaaaacaccttaaattgtcaccgaaatcac 403

Qy     446 atgagttctgggctttcagttatgatgagatggctaggtttgaccttcctgcagtgataa 505
             || | |||||| |||||||  ||||||||||||  | |||||||| ||    || |
Db     404 cagaatactgggccttcagtttggatgagatggctaaatatgaccttccagccacaatca 463

Qy     506 actttatttgcagaaaacgggccaggaaaagatctattatgtcggctattcacagggca 565
           | |||||  |  ||||||| || ||| |     |||| || || ||| | ||||| ||||
Db     464 attttatcatagagaaaactggacagaagcgactctactacgtgggccactcacaaggca 523

Qy     566 ccaccatgggctttattgcattttccaccatgccagagctggctcagaaaatcaaaatgt 625
           ||||||| |  ||||| |||||||| || |  ||||| |||||| | || || || || |
Db     524 ccaccatagcttttatagcattttctacaaacccagaactggctaaaaagattaagatat 583

Qy     626 attttgctttagcacccatagccactgttaagcatgcaaaaagccccgggaccaaatt 683
```

```
Db    584 tttttgcactggctccagttgtcacagttaaatacacccaaagtcctatgaaaaaact 641


RESULT 12
AAF28689
ID   AAF28689 standard; cDNA; 1308 BP.
XX
AC   AAF28689;
XX
DT   05-APR-2001  (first entry)
XX
DE   Human protein HP03372 coding sequence #2.
XX
KW   Human; hydrophobic domain; immune deficiency; autoimmune disorder;
KW   allergy; tissue growth; regeneration; wound healing; burn; tumour;
KW   periodontal disease; thrombolytic condition; haemostatic condition;
KW   infection; ss.
XX
OS   Homo sapiens.
XX
PN   WO200102563-A2.
XX
PD   11-JAN-2001.
XX
PF   16-JUN-2000; 2000WO-JP03943.
XX
PR   02-JUL-1999;   99JP-0188835.
XX
PA   (SAGA ) SAGAMI CHEM RES CENT.
PA   (PROT-) PROTEGENE INC.
XX
PI   Kato S,  Kimura T;
XX
DR   WPI; 2001-071581/08.
DR   P-PSDB; AAB61608.
XX
PT   New human proteins with hydrophobic domains, useful for the treatment
PT   of immune disorders, tumors, allergic conditions, thrombosis and
PT   microbial infection -
XX
PS   Claim 4; Pages 122-124; 153pp; English.
XX
CC   The present invention relates to human proteins (AAB61608-AAB61617) and
CC   their coding sequences (AAF28679-AAF28698). The proteins of the present
CC   invention have hydrophobic domains and can be used for the treatment of
CC   various immune deficiencies and disorders, such as severe combined
CC   immunodeficiency (SCID), multiple sclerosis, rheumatoid arthritis,
CC   autoimmune pulmonary inflammation, graft-versus-host disease and
CC   Guillain-Barre syndrome. The proteins may also be useful in the treatment
CC   of allergic reactions and conditions, such as asthma and in regulation of
CC   haematopoiesis or lymphoid cell deficiencies. The proteins may also have
CC   utility in compositions used for bone, cartilage, tendon and/or nerve
CC   tissue growth or regeneration as well as wound healing and in the
CC   treatment of burns. The proteins may be used in the treatment of
CC   periodontal disease and in other tooth repair processes. Other uses
CC   include treatment of thrombolytic and haemostatic conditions, treatment
CC   or prevention of tumours and inhibiting infection by bacteria, viruses,
CC   fungi and other parasites.
XX
SQ   Sequence 1308 BP; 402 A; 246 C; 264 G; 396 T; 0 other;


  Query Match           22.5%;  Score 286;  DB 22;  Length 1308;
  Best Local Similarity  67.4%;  Pred. No. 1.3e-80;
  Matches  403;  Conservative    0;  Mismatches  195;  Indels    0;  Gaps    0;

Qy     86 gaaatgtgaattcagtacatatgccaactaaagctgtggacccagaagcattcatgaata 145
Db    119 gatctatgtatggttatgacaagaaaggaaacaatgcaaaccctgaagctaatatgaata 178

Qy    146 ttagtgaaatcatccaacatcaaggctatccctgtgaggaatatgaagtcgcaactgaag 205
Db    179 ttagccagattatttcttactggggttatccttatgaagagtatgatgttacaacaaaag 238

Qy    206 atgggtatatccttttctgttaacaggattcctcgaggcctagtgcaacctaagaagacag 265
Db    239 atggttatatccttggaatttataggattccacatggaagaggatgcccagggaggacag 298

Qy    266 gttccaggcctgtggtgttactgcagcatggcctagttggaggtgctagcaactggattt 325
Db    299 ctccaaagcctgctgtgtatttgcagcatggcttaattgcatctgccagtaactggattt 358

Qy    326 ccaacctgcccaacaatagcctgggcttcattctggcagatgctggttttgacgtgtgga 385
Db    359 gcaacctgcccaacaacagtttggctttccttctggcagatagtggttatgacgtgtggt 418

Qy    386 tggggaacagcaggggaaacgcctggtctcgaaaacacaagacactctccatagaccaag 445
Db    419 tggggaacagccgaggaaacacttggtccagaaaacaccttaaattgtcaccgaaatcac 478

Qy    446 atgagttctgggctttcagttatgatgagatggctaggtttgaccttcctgcagtgataa 505
Db    479 cagaatactgggccttcagtttggatgagatggctaaatatgaccttccagccacaatca 538

Qy    506 actttattttgcagaaaacgggccaggaaaagatctattatgtcggctattcacagggca 565
Db    539 attttatcatagagaaaactggacagaagcgactctactacgtgggccactcacaaggca 598

Qy    566 ccaccatgggctttattgcattttccaccatgccagagctggctcagaaaatcaaaatgt 625
Db    599 ccaccatagcttttatagcattttctacaaacccagaactggctaaaaagattaagatat 658

Qy    626 attttgctttagcacccatagccactgttaagcatgcaaaaagccccgggaccaaatt 683
Db    659 tttttgcactggctccagttgtcacagttaaatacacccaaagtcctatgaaaaaact 716


RESULT 13
AAA44349
ID   AAA44349 standard; cDNA; 617 BP.
XX
AC   AAA44349;
XX
DT   21-AUG-2000  (first entry)
XX
DE   Human secreted expressed sequence tag SEQ ID NO:924.
XX
KW   Human; mouse; chicken; rat; secreted expressed sequence tag; sEST;
KW   expressed sequence tag; EST; probe; chemotactic; proliferative;
KW   immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
KW   thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
KW   antiviral; antidiabetic; antiasthmatic; vulnerary; antiparkinsonian;
KW   antiulcer; osteopathic; neuroprotective; nootropic; antipsoriatic;
KW   cerebroprotective; anticonvulsant; antidepressant; gene therapy;
KW   vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
KW   insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
KW   lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
KW   central nervous system disorder; Alzheimer's disease; stroke;
KW   Parkinson's disease; Huntington's disease; coagulation disorder;
KW   haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
KW   tumour; infection; depression; psoriasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200021991-A1.
XX
PD   20-APR-2000.
XX
PF   15-OCT-1999;   99WO-US24206.
XX
PR   15-OCT-1998;   98US-0104436.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Jacobs K,  McCoy JM,  LaVallie ER,  Collins-Racie LA,  Evans C;
```

```
PI   Merberg D,  Treacy M,  Bowman MR;
XX
DR   WPI; 2000-317938/27.
XX
PT   Isolated polynucleotides, and encoded proteins, comprising secreted
PT   expressed sequence tags (sESTs), useful for treating various disorders
PT   such as autoimmune, infectious, and central nervous system disorders -
XX
PS   Claim 1; Page 441; 803pp; English.
XX
CC   AAA43426 to AAA45925 represent specifically claimed secreted expressed
CC   sequence tags (sESTs), isolated from human, mouse, chicken and rat
CC   tissue sources. The sESTs can have a range of activities depending on
CC   the tissues they were isolated from. The activities include:
CC   chemotactic; proliferative; immunomodulatory; haematopoietic;
CC   chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
CC   cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
CC   antiasthmatic; vulnerary; antiulcer; osteopathic; neuroprotective;
CC   nootropic; antiparkinsonian; antipsoriatic; cerebroprotective;
CC   anticonvulsant; and antidepressant. The sESTs can be used for gene
CC   therapy and in vaccines. The sESTs are useful as probes for the
CC   identification and isolation of full-length cDNAs and genomic DNA
CC   molecules which correspond to the sESTs. Proteins encoded by the sESTs
CC   are useful in assays for determining biological activity and raising
CC   antibodies. They may be useful for treatment of autoimmune disorders
CC   (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC   (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
CC   osteoporosis, osteoarthritis, central nervous system disorders
CC   (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC   disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC   disease), tumours, bacterial, fungal or viral infections, depression and
CC   psoriasis. AAA45926 to AAA45931 represent linker variants which are given
CC   in the exemplification of the present invention.
XX
SQ   Sequence 617 BP; 181 A; 125 C; 139 G; 172 T; 0 other;


  Query Match             20.3%;  Score 257.8;  DB 21;  Length 617;
  Best Local Similarity   70.7%;  Pred. No. 7.9e-72;
  Matches  343;  Conservative    0;  Mismatches  142;  Indels    0;  Gaps    0;

Qy    119 ctgtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccct 178
          |||    ||| ||||     ||||| |||||| | || ||     ||   || || ||
Db    133 ctggaagccctgaagtgactatgaacattagtcagatgattacttattggggatacccaa 192

Qy    179 gtgaggaatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctc 238
          ||| ||||||||||| |  |||||||||| ||||| |||   || || || ||||||
Db    193 atgaagaatatgaagttgtgactgaagatggttatattcttgaagtcaatagaattcctt 252

Qy    239 gaggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcc 298
            ||        | |   ||| |||||    || ||||| |||||  ||||||||||
Db    253 atgggaagaaaaattcagggaatacaggccagagacctgttgtgtttttgcagcatggtt 312

Qy    299 tagttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattc 358
          |  ||| |   || |  |||||||||||||||||||| ||||| ||||| | ||||||||
Db    313 tgcttgcatcagccacaaactggatttccaacctgccgaacaacagccttgccttcattc 372

Qy    359 tggcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaa 418
          |||||||||||||||| ||| |||||| |||| |||||||| |||||| ||||| |  |||
Db    373 tggcagatgctggttatgatgtgtggctgggcaacagcagaggaaacacctgggccagaa 432

Qy    419 aacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatgg 478
          | ||  |      ||   |||   || ||| ||||||||||||| | |||||| ||||
Db    433 gaaacttgtactattcaccagattcagttgaattctgggctttcagctttgatgaaatgg 492

Qy    479 ctaggtttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaaaga 538
          |||  | |||||||| ||    ||  |||| ||| |  ||||||| || ||| |  ||
Db    493 ctaaatatgaccttccagccacaatcgacttcattgtaaagaaaactggacagaagcagc 552

Qy    539 tctattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgc 598
          |  | ||||| ||| |||| |||||||||||||| || |||||||| ||||||||||  |
Db    553 tacactatgttggccattcccagggcaccaccattggttttattgccttttccaccaatc 612

Qy    599 cacag 603
            |||
Db    613 tccag 617


RESULT  14
AAA43744
ID   AAA43744 standard; cDNA; 572 BP.
XX
AC   AAA43744;
XX
DT   21-AUG-2000  (first entry)
XX
DE   Mouse secreted expressed sequence tag SEQ ID NO:319.
XX
KW   Human; mouse; chicken; rat; secreted expressed sequence tag; sEST;
KW   expressed sequence tag; EST; probe; chemotactic; proliferative;
KW   immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
KW   thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
KW   antiviral; antidiabetic; antiasthmatic; vulnerary; antiparkinsonian;
KW   antiulcer; osteopathic; neuroprotective; nootropic; antipsoriatic;
KW   cerebroprotective; anticonvulsant; antidepressant; gene therapy;
KW   vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
KW   insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
KW   lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
KW   central nervous system disorder; Alzheimer's disease; stroke;
KW   Parkinson's disease; Huntington's disease; coagulation disorder;
KW   haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
KW   tumour; infection; depression; psoriasis; ss.
XX
OS   Mus musculus.
XX
PN   WO200021991-A1.
XX
PD   20-APR-2000.
XX
PF   15-OCT-1999;   99WO-US24206.
XX
PR   15-OCT-1998;   98US-0104436.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Jacobs K,  McCoy JM,  LaVallie ER,  Collins-Racie LA,  Evans C;
PI   Merberg D,  Treacy M,  Bowman MR;
XX
DR   WPI; 2000-317938/27.
XX
PT   Isolated polynucleotides, and encoded proteins, comprising secreted
PT   expressed sequence tags (sESTs), useful for treating various disorders
PT   such as autoimmune, infectious, and central nervous system disorders -
XX
PS   Claim 1; Page 285-286; 803pp; English.
XX
CC   AAA43426 to AAA45925 represent specifically claimed secreted expressed
CC   sequence tags (sESTs), isolated from human, mouse, chicken and rat
CC   tissue sources. The sESTs can have a range of activities depending on
CC   the tissues they were isolated from. The activities include:
CC   chemotactic; proliferative; immunomodulatory; haematopoietic;
CC   chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
CC   cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
CC   antiasthmatic; vulnerary; antiulcer; osteopathic; neuroprotective;
CC   nootropic; antiparkinsonian; antipsoriatic; cerebroprotective;
CC   anticonvulsant; and antidepressant. The sESTs can be used for gene
CC   therapy and in vaccines. The sESTs are useful as probes for the
CC   identification and isolation of full-length cDNAs and genomic DNA
CC   molecules which correspond to the sESTs. Proteins encoded by the sESTs
CC   are useful in assays for determining biological activity and raising
CC   antibodies. They may be useful for treatment of autoimmune disorders
CC   (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC   (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
CC   osteoporcsis, osteoarthritis, central nervous system disorders
CC   (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC   disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
```

```
CC   disease), tumours, bacterial, fungal or viral infections; depression and
CC   psoriasis. AAA45926 to AAA45931 represent linker variants which are given
CC   in the exemplification of the present invention.
XX
SQ   Sequence 572 BP; 179 A; 109 C; 124 G; 156 T; 4 other;


  Query Match             15.2%;  Score 193.2;  DB 21;  Length 572;
  Best Local Similarity   64.1%;  Pred. No. 2.8e-51;
  Matches  288;  Conservative    0;  Mismatches  161;  Indels    0;  Gaps    0;

Qy     123 ggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtga 182
           | | || |||||   ||||||| |||| || || ||  ||||    |  |||||  | ||
Db     123 gaatcctgaagctcacatgaatgttagcgagattattaaacactgggattatccaagcga 182

Qy     183 ggaatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctcgagg 242
             ||||||| || |  ||||| ||||| || || ||| |  ||||| | |||||||  ||
Db     183 anaatatgaggttgtgactgatgatggttacattcttccaattaaccgaattcctcatgg 242

Qy     243 cctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagt 302
               |  |||| |   |||   | | |   || || ||    |||||||| |  |
Db     243 gaagaacaatgctaatagttcagccccaaagatggtagtattttgtcagcatggcttgct 302

Qy     303 tggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggc 362
           || |    || |     || ||||||| | |||   ||| ||||||| ||||||  | ||
Db     303 tgcaacacctggagcaagggtttccaatccgcctgtcaacagcctggccttcatcntagc 362

Qy     363 agatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaaca 422
           ||||||||| | ||| |||||||||||| | |||||| ||||   |||| |    |||||
Db     363 agatgctgggtatgatgtgtggatgggaagcagcagaggaagtacctgggcaaagaaaca 422

Qy     423 caagacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctag 482
           |  | | |||  |   ||     | || ||||||| ||| |||| |||| | |||   |
Db     423 cgtggccctcaacccngattctaaagaattctgggattttagttttgatcaaatgataaa 482

Qy     483 gtttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatcta 542
           | |||||||||| ||    || || || ||| || | ||||| || ||  |  |||| ||
Db     483 atatgaccttccagctaccattaatttcattctggataaaacaggacaaaagcagattta 542

Qy     543 ttatgtcggctattcacagggcaccacca 571
           ||  | |||||||| || || | | | |
Db     543 ctacattggccattctcaaggaaactcga 571


RESULT  15
AAZ34958
ID   AAZ34958 standard; cDNA; 1718 BP.
XX
AC   AAZ34958;
XX
DT   28-FEB-2000  (first entry)
XX
DE   Soybean acid triacylglycerol lipase cDNA.
XX
KW   Triacylglycerol lipase; soybean; fatty acid; seed oil;
KW   vegetable oil; transgenic plant; ss.
XX
OS   Glycine max.
XX
FH   Key             Location/Qualifiers
FT   CDS             184..1416
FT                   /*tag=  a
XX
PN   WO9955883-A2.
XX
PD   04-NOV-1999.
XX
PF   29-APR-1999;   99WO-US09280.
XX
PR   30-APR-1998;   98US-0083688.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Cahoon EB,  Cahoon RE,  Kinney AJ,  Rafalski JA;
XX
DR   WPI; 2000-062036/05.
DR   P-PSDB; AAY32309.
XX
PT   Novel plant triacylglycerol lipase polynucleotides used to alter the
PT   level of the enzyme in transgenic plants -
XX
PS   Claim 2; Page 48; 65pp; English.
XX
CC   This is the nucleotide sequence of a contig assembled from the
CC   entire cDNA insert in clone ssl.pk0022.a1 and a portion of the
CC   cDNA insert in clone sdp3c.pk004.n3 encoding the entire soybean
CC   acid triacylglycerol lipase (TAGL) (see AAY32309). The clones were
CC   isolated from soybean developing pod and seedling (5-10 day post
CC   germination) cDNA libraries. Novel acid and neutral TAGL polypeptides
CC   (see AAY32301-17) and polynucleotides (see AAZ34950-66) from corn,
CC   Catalpa, rice, soybean and wheat tissues are disclosed. The enzymes
CC   may be prepared recombinantly and used to raise antibodies, which
CC   are used for detecting the enzymes in situ in cells or in vitro in
CC   cell extracts. The polynucleotides may be used to create transgenic
CC   plants in which the TAGL levels are present at higher or lower levels
CC   than normal, or in cell types or developmental processes where they are
CC   not normally found. This would alter the level of triacylglycerol and
CC   cholesteryl esters found in those cells. Accumulation of fatty acids
CC   with unusual structures may be a positive phenotype in plants used for
CC   foods. In addition, it may be desirable to eliminate expression of TAGL
CC   genes for certain applications. TAGL enzymes may also be useful for the
CC   processing of plant seed oils and for the development of novel seed
CC   oils. The TAGL enzymes can also be used as targets to facilitate the
CC   design and/or identification of inhibitors of those enzymes that may be
CC   useful as herbicides. This is desirable because inhibition of the
CC   activity of either of the enzymes could lead to an inhibition of plant
CC   growth. The polynucleotides also serve as a source of probes and
CC   primers, which are useful for genetic mapping, as markers for traits
CC   linked to those genes, and to isolate homologous sequences from other
CC   species.
XX
SQ   Sequence 1718 BP; 492 A; 353 C; 344 G; 529 T; 0 other;


  Query Match             7.9%;  Score 100;  DB 21;  Length 1718;
  Best Local Similarity   55.4%;  Pred. No. 2.5e-21;
  Matches  237;  Conservative    0;  Mismatches  185;  Indels    6;  Gaps    2;

Qy     166 caaggctatccctgtgaggaatatgaagtcgcaactgaagatgggtatatcctttctgtt 225
           || || ||    ||| | ||| | || ||  |||||||  |||||  || || ||     |
Db     337 catggatacaagtgtcaagaacacgaggttacaactgatgatggttacattctgagcctg 396

Qy     226 aacaggattcctcgaggcctagtgcaacctaagaagacaggttccag---gcctgtggtg 282
           | ||||| ||   ||| | ||   ||   |    ||  || | ||   ||     |||
Db     397 caaaggatcccagaaggtcgaggtaaaagcagtgggagtgggacaaggaagcaaccagtg 456

Qy     283 ttactgcagcatggcctagttggaggtgctagcaactggatttccaacctgcccaacaat 342
           |  | || |||||  |  ||| || || ||  |  ||| ||   ||||  ||  |  |
Db     457 gttatacaacatggagttcttgtagatggtatgacatggcttctaaacccaccagagcaa 516

Qy     343 agcctgggcttcattctggcagatgctggttttgacgtgtggatggggaacagcagggga 402
             |||   || ||| | || |||  ||| ||||||||||||||  |  ||||  || |||
Db     517 gatctgccgttgattttagctgataatggatttgacgtgtggattgcaaacacaagagga 576

Qy     403 aacgcctggtctcgaaaacacaagacactctccatagaccaagatgagttctgggctttc 462
           | |   |    |||   |||||   || |   |          | |  |  |||  ||
Db     577 accagatatagtcgccgacacatctcattggacccctctagccaggcctattggaattgg 636

Qy     463 agttatgatgagatggctaggtttgaccttcctgcagtgataaactttattttgcagaaa 522
             || |||||  | |     | |||  | ||||| ||| | || | | | ||     ||
Db     637 tcttgggatgaacttgtctcctatgatttccctgcggtgtttaattatgtgttcagccaa 696

Qy     523 acgggccaggaaaagatctattatgtcggctattcacagggcaccaccatgggctttatt 582
           ||||  |||   |||||| |||| || ||| |||||  ||| ||     | |  ||
Db     697 acggggcag---aagatcaattacgttggccattcattgggaactttggtagctttggca 753
```

```
Qy     583 gcattttc 590
             | || ||
Db     754 tccttctc 761
```

Search completed: August 24, 2001, 22:42:12
Job time: 7433 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: August 24, 2001, 21:22:40 ; Search time 2518.61 Seconds
(without alignments)
4762.814 Million cell updates/sec

Title: US-09-333-159-46
Perfect score: 1269
Sequence: 1 atgttggaaaccttgtcaag.....gacggtgtgaggccgtattg 1269

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_esthum29:*
177: em_esthum30:*
178: em_esthum31:*
179: em_esthum32:*
180: em_esthum33:*
181: em_estom3:*
182: em_estpl11:*
183: em_estro21:*
184: em_estro22:*
185: em_estro23:*
186: em_htc:*
187: gb_est107:*
188: gb_est108:*
189: gb_est109:*
190: gb_est110:*
191: gb_est111:*
192: gb_htc:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_other:*
207: em_gss_pln1:*
208: em_gss_pln2:*
209: em_gss_pro:*
210: em_gss_rod1:*
211: em_gss_rod2:*
212: em_gss_rod3:*
213: em_gss_rod4:*
214: em_gss_rod5:*
215: em_gss_vrt1:*
216: em_gss_vrt2:*
217: em_gss_vrt3:*
218: gb_gss1:*
219: gb_gss2:*
220: gb_gss3:*
221: gb_gss4:*
222: gb_gss5:*
223: gb_gss6:*
224: gb_gss7:*
225: gb_gss8:*
226: gb_gss9:*
227: gb_gss10:*
228: gb_gss11:*
229: gb_gss12:*
230: gb_gss13:*
231: gb_gss14:*
232: gb_gss15:*
233: gb_gss16:*
234: gb_gss17:*
235: gb_gss18:*
236: gb_gss19:*
237: gb_gss20:*
238: gb_gss21:*
239: gb_gss22:*
240: gb_gss23:*
241: gb_gss24:*
242: gb_gss25:*
243: gb_gss26:*
244: gb_gss27:*
245: gb_gss28:*
246: gb_gss29:*
247: gb_gss30:*
248: gb_gss31:*
249: gb_gss32:*
250: gb_gss33:*
251: gb_gss34:*
252: em_gss_inv4:*
253: em_gss_rod6:*
254: em_gss_rod7:*
255: em_gss_rod8:*
256: gb_gss35:*
257: gb_gss36:*
258: gb_gss37:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	618.2	48.7	2927	192	AK019504	AK019504 Mus muscu
2	505	39.8	954	144	BF135102	BF135102 601780142
3	497.6	39.2	1079	145	BF141567	BF141567 601788654
4	471.8	37.2	669	145	BF143546	BF143546 601790896
5	461.8	36.4	1350	192	AK010093	AK010093 Mus muscu
6	460.2	36.3	1347	192	AK010116	AK010116 Mus muscu
7	460.2	36.3	1349	192	AK009300	AK009300 Mus muscu
8	460.2	36.3	1350	192	AK010139	AK010139 Mus muscu
9	458.6	36.1	1340	192	AK010124	AK010124 Mus muscu
10	458.6	36.1	1343	192	AK009479	AK009479 Mus muscu
11	458.6	36.1	1345	192	AK009413	AK009413 Mus muscu
12	458.6	36.1	1345	192	AK009473	AK009473 Mus muscu
13	458.6	36.1	1346	192	AK009359	AK009359 Mus muscu
14	458.6	36.1	1346	192	AK009523	AK009523 Mus muscu
15	458.6	36.1	1346	192	AK009729	AK009729 Mus muscu
16	458.6	36.1	1348	192	AK010058	AK010058 Mus muscu
17	458.6	36.1	1349	192	AK009474	AK009474 Mus muscu
18	458.6	36.1	1349	192	AK009546	AK009546 Mus muscu
19	458.6	36.1	1349	192	AK009571	AK009571 Mus muscu
20	458.6	36.1	1349	192	AK009573	AK009573 Mus muscu
21	458.6	36.1	1349	192	AK010148	AK010148 Mus muscu
22	458.6	36.1	1350	192	AK009459	AK009459 Mus muscu
23	458.6	36.1	1350	192	AK009525	AK009525 Mus muscu
24	458.6	36.1	1350	192	AK009773	AK009773 Mus muscu
25	458.6	36.1	1350	192	AK010061	AK010061 Mus muscu
26	458.6	36.1	1350	192	AK010125	AK010125 Mus muscu
27	458.6	36.1	1350	192	AK010203	AK010203 Mus muscu
28	458.6	36.1	1351	192	AK009428	AK009428 Mus muscu
29	458.6	36.1	1352	192	AK009431	AK009431 Mus muscu
30	458.6	36.1	1356	192	AK010019	AK010019 Mus muscu
31	458.6	36.1	1356	192	AK010035	AK010035 Mus muscu
32	457.6	36.1	1346	192	AK009437	AK009437 Mus muscu
33	457.6	36.1	1347	192	AK010103	AK010103 Mus muscu
34	457.6	36.1	1349	192	AK010026	AK010026 Mus muscu
35	457	36.0	1346	192	AK009560	AK009560 Mus muscu
36	457	36.0	1347	192	AK010236	AK010236 Mus muscu
37	457	36.0	1348	192	AK010231	AK010231 Mus muscu
38	457	36.0	1349	192	AK009875	AK009875 Mus muscu
39	457	36.0	1350	192	AK009537	AK009537 Mus muscu
40	456	35.9	1349	192	AK010106	AK010106 Mus muscu
41	455.4	35.9	1345	192	AK009544	AK009544 Mus muscu
42	455.4	35.9	1349	192	AK009990	AK009990 Mus muscu
43	453.8	35.8	1346	192	AK010173	AK010173 Mus muscu
44	449.8	35.4	1407	192	AK009923	AK009923 Mus muscu
45	448.2	35.3	1350	192	AK010123	AK010123 Mus muscu

ALIGNMENTS

RESULT 1
AK019504
LOCUS AK019504 2927 bp mRNA HTC 08-FEB-2001
DEFINITION Mus musculus 0 day neonate skin cDNA, RIKEN full-length enriched library, clone:4632427C23, full insert sequence.
ACCESSION AK019504
VERSION AK019504.1 GI:12859754
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) 0 day neonate skin cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library clone:4632427C23.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (sites)
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Methods Enzymol. 303, 19-44 (1999)
REFERENCE 2 (sites)
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
REFERENCE 3 (sites)
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Kikuchi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Okazaki,Y., Muramatsu,M., Inoue,Y. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
REFERENCE 4 (sites)
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5 (bases 1 to 2927)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Segabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Gemomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. cDNA of size comprised between 3 and 7 kb was selected before cloning. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B.
FEATURES Location/Qualifiers
source 1. .2927
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="MGD:MGI:1912763"
/db_xref="MGD:MGI:1926003"
/clone="4632427C23"
/tissue_type="skin"
/clone_lib="RIKEN full-length enriched mouse cDNA library"

```
                     /dev_stage="0 day neonate"
     CDS             153. .662
                     /note="putative"
                     /codon_start=1
                     /protein_id="BAB31766.1"
                     /db_xref="GI:12859755"
                     /translation="MSEILSRVWTVSHRVEIWLLILVAYLLQRNVNSGHLPTKAADPE
                     AFMNVSEIIKHKGYPSEEYEVATEDGYILSVNRIPRGQTRLKKEGSRPVVLLQHGLLG
                     DASNWISNLPNNSLGFILADAGFDVWMGNSRGNTWSRKHKTLSIDQDEFWAFRYKNTL
                     RVEVDTDMP"
     polyA_signal    2908. .2913
                     /note="putative"
     polyA_site      2927
                     /note="putative"
BASE COUNT      837 a    581 c    664 g    845 t
ORIGIN


  Query Match             48.7%;  Score 618.2;  DB 192;  Length 2927;
  Best Local Similarity   85.3%;  Pred. No. 8.7e-177;
  Matches  702;  Conservative    0;  Mismatches  118;  Indels    3;  Gaps    1;

Qy     447 tgagttctgggctttcagttatgatgagatggctaggtttgaccttcctgcagtgataaa 506
           ||  || |    ||  |||||||||| ||||||||||||||||||| || |||||||||
Db    1376 TGTCTTTTCCCTTTGTAGTTATGATGAAATGGCTAGGTTTGACCTTCCAGCTGTGATAAA 1435

Qy     507 ctttattttgcagaaaacgggccaggaaaagatctattatgtcggctattcacagggcac 566
           |||||| || |||||||||||||||| ||||| |||||||||||||||| ||||||||||
Db    1436 CTTTATCTTACAGAAAACGGGCCAGAAAAAGGTCTATTATGTCGGCTACTCACAGGGCAC 1495

Qy     567 caccatgggctttattgcattttccaccatgccagagctggctcagaaaatcaaaatgta 626
           |||||||||||||||||||||||||||| |||||||||| ||||| ||||||||||||||
Db    1496 CACCATGGGCTTTATTGCATTTTCCACAATGCCAGAGCTAGCTCATAAAATCAAAATGTA 1555

Qy     627 ttttgctttagcacccatagccactgttaagcatgcaaaaagccccgggaccaaatttttt 686
           |||||| ||||| || |||||||||||||  |||||| ||| || || |||||||||  |
Db    1556 TTTTGCCTTAGCTCCTATAGCCACTGTTAAATATGCAAGAAGTCCTGGTACCAAATTTCT 1615

Qy     687 gttgctgccagatatgatgatcaagggattgtttggcaaaaaagaatttctgtatcagac 746
           | |||||||||||||||||||||||| ||| ||||||| | |||| ||| | || |||||
Db    1616 GCTGCTGCCAGATATGATGATCAAGGTATTATTTGGCAGACAAGAGTTTTTATACCAGAC 1675

Qy     747 cagatttctcagacaacttgttatttacctttgtggccaggtgattcttgatcagatttg 806
           |||||| ||||||| ||| |||||||||||||||||||| |||||||||| || || ||
Db    1676 TAGATTTTTCAGACAGCTTTTTATTTACCTTTGTGGCCAGATGATTCTTGACCAAATCTG 1735

Qy     807 tagtaatatcatgttacttctgggtggattcaacaccaacaatatgaacatgagccgagc 866
           || || ||||| ||||| |||||| || || ||||| ||||||||||||||||||| ||||
Db    1736 CAGCAACATCATCTTACTCCTGGGAGGTTTTAACACAAACAATATGAACATGAGCAGAGC 1795

Qy     867 aagtgtatatgctgcccacactcttgctggaacatctgtgcaaaatattctacactggag 926
           || ||| ||||  ||||| || | ||||||||| |||||||| || ||||| ||||||||
Db    1796 AAATGTGTATGTCGCCCATACACCTGCTGGAACGTCTGTGCAGAACATTCTCCACTGGAG 1855

Qy     927 ccaggcagtgaattctggtgaactccgggcatttgactgggggagtgagaccaaaaatct 986
           |||||| |||||||||| ||||| || || |||||||||| ||||||||||||||||||
Db    1856 CCAGGCCGTGAATTCTGGGGAACTTCGTGCCTTTGACTGGGGAAGTGAGACCAAAAATCA 1915

Qy     987 ggaaaaatgcaatcagccaactcctgtaaggtacagagtcagagatatgacggtccctac 1046
           ||| ||||| |||||||||||||| |||||||| |||  |||||||| |||||| ||
Db    1916 GGAGAAATGTAATCAGCCAACTCCTATAAGGTACAAAGTTCGAGATATGATGGTCCCAAC 1975

Qy    1047 agcaatgtggacaggaggtcaggactggctttcaaatccagaagacgtgaaaatgctgct 1106
           |||||||||||| |||||||| |||||||||||||||||||| || |||||||   | ||
Db    1976 AGCAATGTGGACTGGAGGTCAAGACTGGCTTTCAAATCCAGATGATGTGAAAACATTACT 2035

Qy    1107 ctctgaggtgaccaacctcatctaccataagaatattcctgaatgggctcacgtggattt 1166
           |||||| || |||||||||||||||||| ||||| |||||||| |||||||| ||||||||
Db    2036 TTCTGAAGTAACCAACCTCATCTACCACAAGAACATTCCTGAGTGGGCTCATGTGGATTT 2095

Qy    1167 catctggggtttggatgctcctcaccgtatgtacaatgaaatcatccatctgatgcagca 1226
           |||||||||  |||||| ||||| ||| | |||||||||| || |||||||||| |||
Db    2096 CATCTGGGGGCTGGATGCCCCTCAGCGTGTTTACAATGAAATTATACATCTGATGAAGC- 2154

Qy    1227 ggaggagaccaaccttttcccagggacggtgtgaggccgtattg 1269
           ||||| |||||||| |||||||||   ||   || | ||||
Db    2155 --AGGAGCCCAACCTTCCCCAGGGAACCTGCAGGGTCAAATTG 2195


RESULT   2
BF135102
LOCUS       BF135102      954 bp    mRNA            EST       24-OCT-2000
DEFINITION  601780142F1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:4008131 5',
            mRNA sequence.
ACCESSION   BF135102
VERSION     BF135102.1  GI:10974142
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 954)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by:Incyte Genomics, Inc.
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9242  row: c  column: 12
            High quality sequence stop: 695.
FEATURES             Location/Qualifiers
     source          1. .954
                     /organism="Mus musculus"
                     /strain="CZECH II"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4008131"
                     /clone_lib="NCI_CGAP_Lu30"
                     /tissue_type="tumor, metastatic to mammary"
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; transgenic model WNT-1, expression driven by
                     MMTV-LTR enhancer; Cloned unidirectionally.  Primer: Oligo
                     dT.   Library constructed by Life Technologies.
                     Investigator providing samples: Gilbert Smith, NIH"
BASE COUNT      280 a    222 c    222 g    230 t
ORIGIN


  Query Match             39.8%;  Score 505;  DB 144;  Length 954;
  Best Local Similarity   85.3%;  Pred. No. 1.8e-142;
  Matches  598;  Conservative    0;  Mismatches  100;  Indels    3;  Gaps    3;

Qy     493 cctgcagtgataaacttttattttgcagaaaacgggccaggaaaagatctattatgtcggc 552
           || || |||||||||| || || ||||||||||||||| ||||| ||||||||||||||
Db       1 CCAGCTGTGATAAACTTGATCTTACAGAAAACGGGCCAGAAAAAGGTCTATTATGTCGGC 60

Qy     553 tattcacagggcaccaccatgggctttattgcattttccaccatgccagagctggctcag 612
           || ||||||||||||||||||||||||||||||||||||||||| |||||||||| |||||
Db      61 TACTCACAGGGCACCACCATGGGCTTTATTGCATTTTCCACAATGCCAGAGCTAGCTCAT 120

Qy     613 aaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaaaaagcccc 672
           |||||||||||||||||||| ||||| || ||||||||||||||  |||||| ||| ||
Db     121 AAAATCAAAATGTATTTTGCCTTAGCTCCTATAGCCACTGTTAAATATGCAAGAAGTCCT 180

Qy     673 gggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggcaaaaaagaa 732
           || ||||||||| || ||||||||||||||||||||||||||| ||| ||||||| | ||||
Db     181 GGTACCAAATTTCTGCTGCTGCCAGATATGATGATCAAGGTATTATTTGGCAGACAAGAG 240

Qy     733 tttctgtatcagaccagatttctcagacaacttgttatttacctttgtggccaggtgatt 792
           ||| | || ||||| |||||| |||| || ||| |||||||||| |||||||||| |||||
```

```
Db    241 TTTTTATACCAGACTAGATTTTTCAGGCAGCTTTTTATTTACC-TTGTGGCCAGATGATT 299

Qy    793 cttgatcagatttgtagtaatatcatgttacttctgggtggattcaacaccaacaatatg 852
          || || || || || || || ||||| ||||| ||||| || || ||||| |||||||||
Db    300 CTCGACCAAATCTGCAGCAACATCATCTTACTCCTGGGAGGTTTTAACACAAACAATATG 359

Qy    853 aacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctgtgcaaaat 912
          ||||||||| |||||| ||| |||| ||||| || | |||||||||| |||||||| ||
Db    360 AACATGAGCAGAGCAAATGTGTATGTCGCCCATACACCTGCTGGAACGTCTGTGCAGAAC 419

Qy    913 attctacactggagccaggcagtgaattctggtgaactccgggcatttgactgggggagt 972
          ||||| ||||||||||||| |||||||||| ||||| || || ||||||||||| |||
Db    420 ATTCTCCACTGGAGCCAGGCCGTGAATTCTGGGGAACTTCGTGCCTTTGACTGGGGAAGT 479

Qy    973 gagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagagtcagagat 1032
          |||||||||||| ||| ||||| |||||||||||||| ||||||||| |||  |||||
Db    480 GAGACCAAAAATCAGGAGAAATGTAATCAGCCAACTCCTATAAGGTACAAAGTTCGAGAT 539

Qy   1033 atgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatccagaagac 1092
          |||| |||||| |||||||||||||| |||||||| |||||| ||||||||||||| ||
Db    540 ATGATGGTCCCAACAGCAATGTGGACTGGAGGTCAAGACTGG-TTTCAAATCCAGATGAT 598

Qy   1093 gtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattcctgaatgg 1152
          ||||||    | || ||||| || ||||| |||||||||| ||||| ||||||||   |
Db    599 GTGAAACAATTACTTTCTGAAGTAACCAAACTCATCTACCACAAGAACATTCCTGAGGTG 658

Qy   1153 gctcacgtggatttcatctggggtttggatgctcctcaccg 1193
          | ||| ||||| |||||||||||  ||||||| ||||| ||
Db    659 GGTCATGTGGA-TTCATCTGGGGGGTGGATGCCCCTCAGCG 698


RESULT   3
BF141567
LOCUS       BF141567     1079 bp    mRNA            EST       24-OCT-2000
DEFINITION  601788654F1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:4016413 5',
            mRNA sequence.
ACCESSION   BF141567
VERSION     BF141567.1  GI:10980607
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1079)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by:Incyte Genomics, Inc.
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9263  row: 1  column: 14
            High quality sequence stop: 663.
FEATURES             Location/Qualifiers
     source          1. .1079
                     /organism="Mus musculus"
                     /strain="CZECH II"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4016413"
                     /clone_lib="NCI_CGAP_Lu30"
                     /tissue_type="tumor, metastatic to mammary"
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; transgenic model WNT-1, expression driven by
                     MMTV-LTR enhancer; Cloned unidirectionally.  Primer: Oligo
                     dT.   Library constructed by Life Technologies.
                     Investigator providing samples: Gilbert Smith, NIH"
BASE COUNT      312 a    267 c    263 g    237 t
```

```
ORIGIN


  Query Match           39.2%;  Score 497.6;  DB 145;  Length 1079;
  Best Local Similarity   84.1%;  Pred. No. 3.4e-140;
  Matches  621;  Conservative    0;  Mismatches  109;  Indels    8;  Gaps    5;

Qy    256 aagaagacaggttccaggcctgtggtgttactgcagcatggcctagttggaggtgctagc 315
          ||||||  ||| |||||||| |||||||||||||| ||||| ||  |||| | |||||||
Db     20 AAGAAGGAAGGATCCAGGCCAGTGGTGTTACTGCAACATGGTCTTCTTGGGGATGCTAGC 79

Qy    316 aactggatttccaacctgcccaacaatagcctgggcttcattctggcagatgctggtttt 375
          ||||||||||||||||||||||||||||| |||||||| || ||||||||||||| ||||||
Db     80 AACTGGATTTCCAACCTGCCCAACAACAGCCTGGGTTTTATTCTGGCAGATGCAGGTTTT 139

Qy    376 gacgtgtggatggggaacagcaggggaaacgcc-tggtctcgaaaacacaagacactctc 434
          || |||||||||| |||||||| |||||| || |||||||| || |||||||| ||||||
Db    140 GATGTGTGGATGGGAAACAGCAGAGGAAACACCGTGGTCTCGGAAGCACAAGACCCTCTC 199

Qy    435 catagaccaagatgagttctgggctttcagttatgatgagatggctaggtttgaccttcc 494
          |||||| ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    200 CATAGATCAAGATGAGTTCTGGGCTTTCAGTTATGATGAAATGGCTAGGTTTGACCTTCC 259

Qy    495 tgcagtgataaacttttattttgcagaaaacgggccaggaaaagatctattatgtcggcta 554
          || |||||||||||||| || |||||||||||||| ||||| ||||||||||||||||||
Db    260 AGCTGTGATAAACTTTATCTTACAGAAAACGGGCCAGAAAAAGGTCTATTATGTCGGCTA 319

Qy    555 ttcacagggcaccaccatgggctttattgcattttccaccatgccagagctggctcagaa 614
          |||||||||||||||||||||||||||||||||||||||||||| |||||||||| ||||| ||
Db    320 CTCACAGGGCACCACCATGGGCTTTATTGCATTTTCCACAATGCCAGAGCTAGCTCATAA 379

Qy    615 aatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaaaaagccccgg 674
          |||||||||||||||| || ||||| || |||||||||||||  |||||| ||| || ||
Db    380 AATCAAAATGTATTTGGCCTTAGCTCCTATAGCCACTGTTAAATATGCAAGAAGTCCTGG 439

Qy    675 gaccaaatttttgttgctgccagatatgatgatcaagggattgtttggcaaaaaagaatt 734
          |||||||||  || |||||||||||||||||||||||| ||| ||||||| | |||| ||
Db    440 TACCAAATTTCTGCTGCTGCCAGATATGATGATCAAGGTATTATTTGGCAGACAAGAGTT 499

Qy    735 tctgtatcagaccagatttctcagacaacttgttatttaccctttgtggccaggtgattct 794
          | | || ||||| |||||| |||| || ||| ||||||||||||||| ||| |||||||
Db    500 TTTATACCAGACTAGATTTTTCAGGCAGCTTTTTATTTACCTTTGTGGGCAGATGATTCT 559

Qy    795 tgatcagatttgtagtaatatcatgttacttctgggtggattcaacaccaacaatatgaa 854
          || || || || || || || ||||| ||||| ||||| || || ||||| ||| |||||||
Db    560 CGACCAAATCTGCAGCAACATCATCTTACTCCTGGGAGGTTTTAACACAAACCATATGAA 619

Qy    855 catgagccgagca--agtgtatatgctgcccacactcttgctggaacatctgtg---caa 909
          ||||||| ||||| |     | || ||||| |  | |||||||| ||| |   |   | |
Db    620 CATGAGCAGAGCAACATGTGCTCTGTCGCCCATAAACCTGCTGGAACGTCTTGGTGCCGA 679

Qy    910 aatattctacactggagccagg-cagtgaattctggtg-aactccgggcatttgactggg 967
          |  ||||| |||||||||||||| | |||||||||| | | || || || |||||||| |
Db    680 ACCATTCTCCACTGGAGCCAGGCCGGTGAATTCTGGGGCACCTTCGTGCCTTTGACTGTG 739

Qy    968 ggagtgagaccaaaaatc 985
          ||   | ||| |  || |
Db    740 GGCAAGCGACAACCAAGC 757


RESULT   4
BF143546
LOCUS       BF143546     669 bp    mRNA            EST       24-OCT-2000
DEFINITION  601790896F1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:4021870 5',
            mRNA sequence.
ACCESSION   BF143546
VERSION     BF143546.1  GI:10982586
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 669)
```

```
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by:Incyte Genomics, Inc.
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9277  row: o  column: 23
            High quality sequence stop: 659.
FEATURES             Location/Qualifiers
     source          1. .669
                     /organism="Mus musculus"
                     /strain="C2ECH II"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4021870"
                     /clone_lib="NCI_CGAP_Lu30"
                     /tissue_type="tumor, metastatic to mammary"
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; transgenic model WNT-1, expression driven by
                     MMTV-LTR enhancer; Cloned unidirectionally.  Primer: Oligo
                     dT.   Library constructed by Life Technologies.
                     Investigator providing samples: Gilbert Smith, NIH"
BASE COUNT      195 a    140 c    165 g    169 t
ORIGIN


  Query Match           37.2%;  Score 471.8;  DB 145;  Length 669;
  Best Local Similarity   85.9%;  Pred. No. 2e-132;
  Matches  535;  Conservative    0;  Mismatches   87;  Indels    1;  Gaps    1;

Qy      1 atgttggaaaccttgtcaagacagtggattgtctcacacagaatggaaatgtggcttctg 60
          ||||  |||| ||||||||||  ||||| ||| || |||||| |||| || |||||||||
Db     48 ATGTCAGAAATCTTGTCAAGAGTGTGGACTGTTTCGCACAGAGTGGAGATATGGCTTCTG 107

Qy     61 attctggtggcgtatatgttccagagaaatgtgaattcagtacatatgccaactaaagct 120
          |||||||| || ||| |  |||| |||||||||| || | |||| |||| || ||||||
Db    108 ATTCTGGTAGCATATTTACTCCAAAGAAATGTGAACTCGGGACATTTGCCCACGAAAGCT 167

Qy    121 gtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgt 180
          | ||| |||||||||||||||||| |||| ||||||||| |||| :| || |||||| ||
Db    168 GCGGATCCAGAAGCATTCATGAATGTTAGCGAAATCATCAAACACAAGGGTTATCCCAGT 227

Qy    181 gaggaatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctcga 240
          ||||| |||||||| ||||| |||||||||||| |||||||||| ||||| || ||||||
Db    228 GAGGAGTATGAAGTTGCAACAGAAGATGGGTACATCCTTTCTGTGAACAGAATCCCTCGG 287

Qy    241 ggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggccta 300
          || |     |     ||||||  ||| |||||||| |||||||||||||| ||||| ||
Db    288 GGACAGACACGGTTAAAGAAGGAAGGATCCAGGCCAGTGGTGTTACTGCAACATGGTCTT 347

Qy    301 gttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctg 360
          |||| | ||||||||||||||||||||||||||||||| |||||||| || ||||||
Db    348 CTTGGGGATGCTAGCAACTGGATTTCCAACCTGCCCAACAACAGCCTGGGTTTTATTCTG 407

Qy    361 gcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaa 420
          |||||||| || ||||| |||||||||| |||||||| |||||| ||||||||||| ||
Db    408 GCAGATGCAGG-TTTGATGTGTGGATGGGAAACAGCAGAGGAAACACCTGGTCTCGGAAG 466

Qy    421 cacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatggct 480
          |||||||| |||||||||| |||||||||||||||||||||||||||||||||| ||||||
Db    467 CACAAGACCCTCTCCATAGATCAAGATGAGTTCTGGGCTTTCAGTTATGATGAAATGGCT 526

Qy    481 aggtttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatc 540
          |||||||||||||| || |||||||||||||| || |||||||||||||| ||||| ||
Db    527 AGGTTTGACCTTCCAGCTGTGATAAACTTTATCTTACAGAAAACGGGCCAGAAAAAGGTC 586

Qy    541 tattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgcca 600
          |||||||||||||| |||||||||||||||||||||||||||||||||||||| ||||||
Db    587 TATTATGTCGGCTACTCACAGGGCACCACCATGGGCTTTATTGCATTTTCCACAATGCCA 646

Qy    601 gagctggctcagaaaatcaaaat 623
          ||||| ||||| || ||| ||||
Db    647 GAGCTAGCTCATAACATCCAAAT 669


RESULT   5
AK010093
LOCUS       AK010093     1350 bp    mRNA            HTC       08-FEB-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310067K20, full insert sequence.
ACCESSION   AK010093
VERSION     AK010093.1  GI:12845297
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310067K20.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods Enzymol. 303, 19-44 (1999)
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Kikuchi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T.,
            Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T.,
            Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Okazaki,Y., Muramatsu,M., Inoue,Y. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1350)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
```

```
          Encyclopedia Project of Genome Exploration Research Group in Riken
          Genomic Sciences Center and Genome Science Laboratory in RIKEN.
          Division of Experimental Animal Research in Riken contributed to
          prepare mouse tissues. First strand cDNA was primed with a primer
          [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
          prepared by using trehalose thermo-activated reverse transcriptase
          and subsequently enriched for full-length by cap-trapper. cDNA went
          through one round of normalization to Rot = 5.0 and subtraction to
          Rot = 25.0. Second strand cDNA was prepared with the primer adapter
          of sequence [5'
          GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved
          with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
          Host: SOLR.
FEATURES             Location/Qualifiers
     source          1. .1350
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /db_xref="MGD:MGI:1898392"
                     /db_xref="MGD:MGI:1914967"
                     /clone="2310067K20"
                     /sex="male"
                     /tissue_type="tongue"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             53. .1240
                     /note="putative"
                     /codon_start=1
                     /protein_id="BAB26697.1"
                     /db_xref="GI:12845298"
                     /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                     SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLVASATNWITNLPNNSLA
                     FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEEFWAFSFDEMAKYDLPATIDFIVQKT
                     GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
                     FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV
                     YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                     AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                     AED"
BASE COUNT      396 a    290 c    286 g    378 t
ORIGIN
```

```
  Query Match           36.4%;  Score 461.8;  DB 192;  Length 1350;
  Best Local Similarity   63.8%;  Pred. No. 3e-129;
  Matches  700;  Conservative    0;  Mismatches  397;  Indels    0;  Gaps    0;

Qy     125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
           |||| ||||||  |||||||  |||||  | ||  ||     |   || |||||  ||||||
Db     132 ACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTACTGGGGATATCCAAGTGAGG 191

Qy     185 aatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctcgaggcc 244
           ||||||||||  |  ||||||||||| || || ||    ||  | || ||||||   ||
Db     192 AATATGAAGTTGTTACTGAAGATGGCTACATTCTGGGGGTCTATAGAATTCCTTATGGGA 251

Qy     245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
               |  || |||| |  ||    || |||||||  |   ||||||||||  | ||||
Db     252 AGAAAAATTCTGAGAATATCGGCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGGTTG 311

Qy     305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
           |   || |  |||||||| | || ||||| ||||| |||||||  |||||||| ||||
Db     312 CATCAGCCACAAACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAG 371

Qy     365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
           |||||||  | |||  ||||||  ||||||||||   | ||  ||   | |||||  || ||| |
Db     372 ATGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCCGGAAAAATG 431

Qy     425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
                 ||   ||||  || ||| ||||||||||||||  | ||||||  |||||||  |
Db     432 TATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGGCTAAAT 491

Qy     485 ttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 544
           |||||||||  ||    |||  |||| |||  | ||||||| ||  ||  ||  |||||   | |
Db     492 ATGACCTTCCAGCCACCATAGACTTCATTGTACAGAAAACTGGACAAGAGAAGATACACT 551

Qy     545 atgccggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
           ||||  ||   |  ||  |||||||||||| ||  ||  |||||||| |||||  ||||   ||  |  |
Db     552 ATGTTGGTCACTCTCAGGGCACCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTC 611

Qy     605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
           |||||  |  ||||||||  |  ||  ||  |||  |||||  ||   |  ||  |||||  |||  ||  ||
Db     612 TGGCTAAAAAAAATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAG 671

Qy     665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
           ||||  |||   .|   ||  |||   |   |  ||   |   |   |  ||||||   | |||||  |
Db     672 AAAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGATATTTGGTA 731

Qy     725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccttttgtggcc 784
           |  |||   ||  ||   ||  |  |   ||       |   ||||  ||   |  |  ||   |
Db     732 ACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTTGGTACGGAAGTGTGCTCAC 791

Qy     785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
           ||  |  |  ||  ||||    |  ||  ||  ||      |  ||   |   |  |||||||  |||  |
Db     792 GGGAGCTGCTAGATCTTCTCTGCAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGA 851

Qy     845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
           |  ||   |  ||||  ||       |||  |||    |   ||  |  ||  ||  |||||||||
Db     852 AAAACTTAAATGTGAGTCGCTTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTA 911

Qy     905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
             |||  |   ||  |  ||||||    |||    |   ||  |||||  |  ||  |   |||
Db     912 CTCAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGCCTATAACT 971

Qy     965 ggggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
           |||  ||||        |  ||   ||   |  |  |  |||||||   |||  |||       ||     |
Db     972 GGGGGAGTCCATTACAGAACATGTTACACTACAATCAGAAAACGCCTCCCTACTATGATG 1031

Qy    1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
           |   ||  |||||  ||  ||  |   |||  ||||||   ||  ||  ||  |||    ||   |   |||
Db    1032 TGTCAGCCATGACCGTGCCAATTGCAGTGTGGAACGGTGGCCATGACATCCTGGCTGATC 1091

Qy    1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
           |   |||  ||   |||||||||  |   |   |  ||||||   |  |||||||||  |  ||||
Db    1092 CCCAAGATGTCGCAATGCTGCTTCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTC 1151

Qy    1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
           |   |    ||||  ||||  ||||||||||   |||||||  |||||    |  |||||||
Db    1152 TTCCCTACAATCACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATG 1211

Qy    1205 aaatcatccatctgatg 1221
           |  ||  |    |||||
Db    1212 AGATAGTTACCATGATG 1228
```

```
RESULT   6
AK010116
LOCUS       AK010116     1347 bp    mRNA            HTC       08-FEB-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310068O05, full insert sequence.
ACCESSION   AK010116
VERSION     AK010116.1  GI:12845334
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310068O05.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods Enzymol. 303, 19-44 (1999)
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
```

```
JOURNAL    Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE    20499374
REFERENCE  3  (sites)
AUTHORS    Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
           Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
           Kikuchi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T.,
           Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T.,
           Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
           Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
           Matsuura,S., Okazaki,Y., Muramatsu,M., Inoue,Y. and Hayashizaki,Y.
TITLE      RIKEN integrated sequence analysis (RISA) system--384-format
           sequencing pipeline with 384 multicapillary sequencer
JOURNAL    Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE    20530913
REFERENCE  4  (sites)
AUTHORS    The RIKEN Genome Exploration Research Group Phase II Team and
           FANTOM Consortium.
TITLE      Functional annotation of a full-length mouse cDNA collection
JOURNAL    Nature 409, 685-690 (2001)
REFERENCE  5  (bases 1 to 1347)
AUTHORS    Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
           Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
           Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
           Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
           Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
           Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
           Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
           Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
           Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
           Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
           Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE      Direct Submission
JOURNAL    Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
           Physical and Chemical Research (RIKEN), Laboratory for Genome
           Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
           RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
           Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
           URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
           Fax:81-45-503-9216)
COMMENT    Please visit our web site (http://genome.gsc.riken.go.jp/) for
           further details.
           cDNA library was prepared and sequenced in Mouse Genome
           Encyclopedia Project of Genome Exploration Research Group in Riken
           Genomic Sciences Center and Genome Science Laboratory in RIKEN.
           Division of Experimental Animal Research in Riken contributed to
           prepare mouse tissues. First strand cDNA was primed with a primer
           [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
           prepared by using trehalose thermo-activated reverse transcriptase
           and subsequently enriched for full-length by cap-trapper. cDNA went
           through one round of normalization to Rot = 5.0 and subtraction to
           Rot = 25.0. Second strand cDNA was prepared with the primer adapter
           of sequence [5'
           GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved
           with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
           Host: SOLR.
FEATURES            Location/Qualifiers
     source         1..1347
                    /organism="Mus musculus"
                    /strain="C57BL/6J"
                    /db_xref="taxon:10090"
                    /db_xref="MGD:MGI:1895801"
                    /db_xref="MGD:MGI:1914967"
                    /clone="2310068005"
                    /sex="male"
                    /tissue_type="tongue"
                    /clone_lib="RIKEN full-length enriched mouse cDNA library"
                    /dev_stage="adult"
     CDS            51..1238
                    /note="putative"
                    /codon_start=1
                    /protein_id="BAB26711.1"
                    /db_xref="GI:12845335"
                    /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                    SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
                    FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                    GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
                    FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFGV
                    YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                    AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                    AED"
BASE COUNT     393 a    290 c    284 g    380 t
ORIGIN

  Query Match            36.3%;  Score 460.2;  DB 192;  Length 1347;
  Best Local Similarity  63.7%;  Pred. No. 9.2e-129;
  Matches  699;  Conservative    0;  Mismatches  398;  Indels    0;  Gaps    0;

Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
Db    130 ACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTACTGGGGATATCCAAGTGAGG 189

Qy    185 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 244
Db    190 AATATGAAGTTGTTACTGAAGATGGCTACATTCTGGGGGTCTATAGAATTCCTTATGGGA 249

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
Db    250 AGAAAAATTCTGAGAATATCGGCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTG 309

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
Db    310 CATCAGCCACAAACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAG 369

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
Db    370 ATGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCCGGAAAAATG 429

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
Db    430 TATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGGCTAAAT 489

Qy    485 ttgaccttcctgcagtgataaactttatttgcagaaaacgggccaggaaaagatctatt 544
Db    490 ATGACCTTCCAGCCACCATAGACTTCATTGTACAGAAAACTGGACAAGAGAAGATACACT 549

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
Db    550 ATGTTGGTCACTCTCAGGGCACCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTC 609

Qy    605 tggctcagaaaatcaaaatgtatttgctttagcacccatagccactgttaagcatgcaa 664
Db    610 TGGCTAAAAAAATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAG 669

Qy    665 aaagccccgggaccaaattttgttgctgccagatatgatgatcaagggattgtttggca 724
Db    670 AAAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGATATTTGGTA 729

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
Db    730 ACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTTGGTACGGAAGTGTGCTCAC 789

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
Db    790 GGGAGCTGCTAGATCTTCTCTGCAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGA 849

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
Db    850 AAAACTTAAATGTGAGTCGCTTTGGTGTGTATCTAGGGCATAATCCAGCAGGAACATCTA 909

Qy    905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
Db    910 CTCAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGCCTATAACT 969

Qy    965 ggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
```

```
Db    970 GGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAACGCCTCCCTACTATGATG 1029

Qy   1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
          |   ||  ||||| || || |  ||| ||||||  || || || |||   ||  |  |||
Db   1030 TGTCAGCCATGACCGTGCCAATTGCAGTGTGGAACGGTGGCCATGACATCCTGGCTGATC 1089

Qy   1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
          |  |||| ||   |||||||||  |  |  |  |||||||  | |||||||||  | ||||
Db   1090 CCCAAGATGTCGCAATGCTGCTTCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTC 1149

Qy   1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
           |   |    |||| |||| ||||||||||   ||||||| |||||     | |||||||
Db   1150 TTCCCTACAATCACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATG 1209

Qy   1205 aaatcatccatctgatg 1221
          | ||  |     |||||
Db   1210 AGATAGTTACCATGATG 1226


RESULT   7
AK009300
LOCUS       AK009300     1349 bp    mRNA            HTC       08-FEB-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310011G18, full insert sequence.
ACCESSION   AK009300
VERSION     AK009300.1  GI:12844011
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310011G18.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods Enzymol. 303, 19-44 (1999)
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Kikuchi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T.,
            Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T.,
            Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Okazaki,Y., Muramatsu,M., Inoue,Y. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1349)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 25.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES              Location/Qualifiers
     source           1. .1349
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /db_xref="MGD:MGI:1895751"
                      /db_xref="MGD:MGI:1914967"
                      /clone="2310011G18"
                      /sex="male"
                      /tissue_type="tongue"
                      /clone_lib="RIKEN full-length enriched mouse cDNA library"
                      /dev_stage="adult"
     CDS              53. .1240
                      /note="putative"
                      /codon_start=1
                      /protein_id="BAB26201.1"
                      /db_xref="GI:12844012"
                      /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                      SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
                      FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                      GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
                      FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFGV
                      YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                      AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                      AED"
BASE COUNT      395 a    290 c    284 g    380 t
ORIGIN


  Query Match             36.3%;  Score 460.2;  DB 192;  Length 1349;
  Best Local Similarity   63.7%;  Pred. No. 9.2e-129;
  Matches  699;  Conservative    0;  Mismatches  398;  Indels    0;  Gaps    0;

Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
          |||| ||||||  |||||||  |||||  | || ||  .   |    || |||||  ||||||
Db    132 ACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTACTGGGGATATCCAAGTGAGG 191

Qy    185 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 244
          ||||||||||  |  ||||||||||| || || ||    ||  | || ||||||   ||
Db    192 AATATGAAGTTGTTACTGAAGATGGCTACATTCTGGGGGTCTATAGAATTCCTTATGGGA 251

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
              |  || |||| |  ||    || |||||||  | . ||||||||||  |  |||
Db    252 AGAAAAATTCTGAGAATATCGGCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTG 311

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
           |   || |  |||||||||  | || |||||  |||||  |||||||  |||||||||  ||||
Db    312 CATCAGCCACAAACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAG 371
```

```
Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
          ||||||| | ||| ||||||| |||||||||| | || ||  | ||||| || ||| |
Db    372 ATGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCCGGAAAAATG 431

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
          ||   ||||  || ||| |||||||||||||| | |||||| |||||||  |
Db    432 TATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGGCTAAAT 491

Qy    485 ttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaaagatctatt 544
          |||||||||  ||    ||| |||| ||| | |||||||| || || || |||||  | |
Db    492 ATGACCTTCCAGCCACCATAGACTTCATTGTACAGAAAACTGGACAAGAGAAGATACACT 551

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
          |||| || | || ||||||||||| || || |||||||| ||||| |||| || |  |
Db    552 ATGTTGGTCACTCTCAGGGCACCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTC 611

Qy    605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
          ||||| | |||||||| | || || ||| ||||| ||  |  || ||||| ||| || ||
Db    612 TGGCTAAAAAAATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAG 671

Qy    665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
          |||| |||   |  ||  |||  |  | ||  |  |  |  ||||||  | |||||  |
Db    672 AAAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGATATTTGGTA 731

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
          | |||  ||  ||  || | |  ||      |  |||| ||  |  | ||    |
Db    732 ACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTTGGTACGGAAGTGTGCTCAC 791

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
          || | | || ||||   | || || ||     | ||  |  |  |||||||  |||  |
Db    792 GGGAGCTGCTAGATCTTCTCTGCAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGA 851

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
          | ||  | ||  |||| ||     |||| |||   |  || | ||  || |||||||||
Db    852 AAAACTTAAATGTGAGTCGCTTTGGTGTGTATCTAGGGCATAATCCAGCAGGAACATCTA 911

Qy    905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
          |||  |  || | ||||||   |||    |  || |||||  | || |  || | | |||
Db    912 CTCAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGCCTATAACT 971

Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
          ||||  |||       | ||  ||  | | | |||||||  |||  |||       ||    |
Db    972 GGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAACGCCTCCCTACTATGATG 1031

Qy   1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
          |   ||  ||||| || || |  ||| ||||||  || || ||  |||   ||. |  |||
Db   1032 TGTCAGCCATGACCGTGCCAATTGCAGTGTGGAACGGTGGCCATGACATCCTGGCTGATC 1091

Qy   1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
          |  |||| ||   |||||||||  |  |  |  |||||||  | |||||||||| | ||||
Db   1092 CCCAAGATGTCGCAATGCTGCTTCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTC 1151

Qy   1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
          |   |    |||| |||| ||||||||||    |||||||  |||||     | |||||||
Db   1152 TTCCCTACAATCACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATG 1211

Qy   1205 aaatcatccatctgatg 1221
          | ||  |     |||||
Db   1212 AGATAGTTACCATGATG 1228
```

```
RESULT   8
AK010139
LOCUS       AK010139     1350 bp    mRNA            HTC       08-FEB-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310069P19, full insert sequence.
ACCESSION   AK010139
VERSION     AK010139.1  GI:12845371
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310069P19.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods Enzymol. 303, 19-44 (1999)
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Kikuchi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T.,
            Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T.,
            Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Okazaki,Y., Muramatsu,M., Inoue,Y. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1350)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 25.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES              Location/Qualifiers
     source           1. .1350
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /db_xref="MGD:MGI:1898377"
                      /db_xref="MGD:MGI:1914967"
```

```
                     /clone="2310069P19"
                     /sex="male"
                     /tissue_type="tongue"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             53. .1240
                     /note="putative"
                     /codon_start=1
                     /protein_id="BAB26725.1"
                     /db_xref="GI:12845372"
                     /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                     SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
                     FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                     GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPIATVKYTESPFKKISLIPK
                     FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV
                     YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                     AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                     AED"
BASE COUNT      398 a    289 c    283 g    380 t
ORIGIN
```

Query Match 36.3%; Score 460.2; DB 192; Length 1350;
Best Local Similarity 63.7%; Pred. No. 9.2e-129;
Matches 699; Conservative 0; Mismatches 398; Indels 0; Gaps 0;

```
Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
Db    132 ACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTACTGGGGATATCCAAGTGAGG 191

Qy    185 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 244
Db    192 AATATGAAGTTGTTACTGAAGATGGCTACATTCTGGGGGTCTATAGAATTCCTTATGGGA 251

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
Db    252 AGAAAAATTCTGAGAATATCGGCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTG 311

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
Db    312 CATCAGCCACAAACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAG 371

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
Db    372 ATGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCCGGAAAAATG 431

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
Db    432 TATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGGCTAAAT 491

Qy    485 ttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 544
Db    492 ATGACCTTCCAGCCACCATAGACTTCATTGTACAGAAAACTGGACAAGAGAAGATACACT 551

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
Db    552 ATGTTGGTCACTCTCAGGGCACCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTC 611

Qy    605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
Db    612 TGGCTAAAAAAATCAAGAGGTTTTATGCATTAGCTCCAATTGCTACTGTGAAGTATACAG 671

Qy    665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
Db    672 AAAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGATATTTGGTA 731

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
Db    732 ACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTTGGTACGGAAGTGTGCTCAC 791

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
Db    792 GGGAGCTGCTAGATCTTCTCTGCAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGA 851
```

```
Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
Db    852 AAAACTTAAATGTGAGTCGCTTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTA 911

Qy    905 tgcaaaatattctacactggagccaggcagtgaatttctggtgaactccgggcatttgact 964
Db    912 CTCAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGCCTATAACT 971

Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
Db    972 GGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAACGCCTCCCTACTATGATG 1031

Qy   1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
Db   1032 TGTCAGCCATGACCGTGCCAATTGCAGTGTGGAACGGTGGCCATGACATCCTGGCTGATC 1091

Qy   1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
Db   1092 CCCAAGATGTCGCAATGCTGCTTCCCAAACTCCCCAACCTTTTGTACCATAAGGAGATTC 1151

Qy   1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
Db   1152 TTCCCTACAATCACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATG 1211

Qy   1205 aaatcatccatctgatg 1221
Db   1212 AGATAGTTACCATGATG 1228
```

```
RESULT   9
AK010124
LOCUS       AK010124     1340 bp    mRNA            HTC       08-FEB-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310069D09, full insert sequence.
ACCESSION   AK010124
VERSION     AK010124.1  GI:12845346
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310069D09.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Seiurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods Enzymol. 303, 19-44 (1999)
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Kikuchi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T.,
            Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T.,
            Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Okazaki,Y., Muramatsu,M., Inoue,Y. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1340)
```

```
AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
          Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
          Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
          Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
          Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
          Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
          Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
          Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
          Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
          Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
          Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE     Direct Submission
JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
          Physical and Chemical Research (RIKEN), Laboratory for Genome
          Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
          RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
          Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
          URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
          Fax:81-45-503-9216)
COMMENT   Please visit our web site (http://genome.gsc.riken.go.jp/) for
          further details.
          cDNA library was prepared and sequenced in Mouse Genome
          Encyclopedia Project of Genome Exploration Research Group in Riken
          Genomic Sciences Center and Genome Science Laboratory in RIKEN.
          Division of Experimental Animal Research in Riken contributed to
          prepare mouse tissues. First strand cDNA was primed with a primer
          [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTVN 3'], cDNA was
          prepared by using trehalose thermo-activated reverse transcriptase
          and subsequently enriched for full-length by cap-trapper. cDNA went
          through one round of normalization to Rot = 5.0 and subtraction to
          Rot = 25.0. Second strand cDNA was prepared with the primer adapter
          of sequence [5'
          GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved
          with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
          Host: SOLR.
FEATURES             Location/Qualifiers
     source          1. .1340
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /db_xref="MGD:MGI:1895811"
                     /db_xref="MGD:MGI:1914967"
                     /clone="2310069D09"
                     /sex="male"
                     /tissue_type="tongue"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             47. .1234
                     /note="putative"
                     /codon_start=1
                     /protein_id="BAB26715.1"
                     /db_xref="GI:12845347"
                     /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                     SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
                     FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                     GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
                     FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV
                     YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                     AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                     AED"
BASE COUNT      393 a    287 c    282 g    378 t
ORIGIN

  Query Match           36.1%;  Score 458.6;  DB 192;  Length 1340;
  Best Local Similarity  63.6%;  Pred. No. 2.8e-128;
  Matches  698;  Conservative    0;  Mismatches  399;  Indels    0;  Gaps    0;

Qy     125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
           |||| ||||||  |||||||  |||||  | || ||     |    || |||||  ||||||
Db     126 ACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTACTGGGGATATCCAAGTGAGG 185

Qy     185 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 244
           |||||||||| |  |||||||||| || || ||    || | || ||||||   ||
Db     186 AATATGAAGTTGTTACTGAAGATGGCTACATTCTGGGGGTCTATAGAATTCCTTATGGGA 245

Qy     245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
           |  || |||| |  ||   || |||||||  |  ||||||||||  |  |||
Db     246 AGAAAAATTCTGAGAATATCGGCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTG 305

Qy     305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
           |   || |  |||||||||  | || |||||  |||||  |||||||  |||||||| ||||
Db     306 CATCAGCCACAAACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAG 365

Qy     365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
           |||||||  | |||  ||||||  ||||||||||   | || ||  | |||||  || ||| |
Db     366 ATGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCCGGAAAAATG 425

Qy     425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
           |  ||  ||||  || ||| ||||||||||||||  | ||||||  |||||||  |
Db     426 TATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGGCTAAAT 485

Qy     485 ttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaaagatctatt 544
           |||||||||  ||    ||| |||| ||| | |||||||| || || || |||||  | |
Db     486 ATGACCTTCCAGCCACCATAGACTTCATTGTACAGAAAACTGGACAAGAGAAGATACACT 545

Qy     545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
           ||||  ||  | ||  |||||||||||  || || |||||||  |||||  ||||  || |  |
Db     546 ATGTTGGTCACTCTCAGGGCACCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTC 605

Qy     605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
           |||||  | |||||||| | ||  ||  |||  |||||  ||   |  ||  |||||  ||| ||  ||
Db     606 TGGCTAAAAAAAATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAG 665

Qy     665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
           ||||  |||   |  ||  |||   |  | ||   |  |  |  |  |  ||||||   | ||||  |
Db     666 AAAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGATATTTGGTA 725

Qy     725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccttttgtggcc 784
           | |||   ||  ||   ||  | |   ||    |  ||||  ||   |  | ||    |
Db     726 ACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTTGGTACGGAAGTGTGCTCAC 785

Qy     785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
           ||  | | ||  ||||   | ||  ||  ||      |  ||   |  |  |||||||  |||  |
Db     786 GGGAGCTGCTAGATCTTCTCTGCAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGA 845

Qy     845 acaatatgaaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
           | ||   | ||  ||||  ||      |||  |||   |  ||  |  ||  || |||||||||
Db     846 AAAACTTAAATGTGAGTCGCTTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTA 905

Qy     905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
           |||  |  || | ||||||   |||   |  || |||||  | ||  |  || | | |||
Db     906 CTCAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGCCTATAACT 965

Qy     965 ggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
           ||||  |||       | ||  ||  | | | |||||||  |||  |||      ||   |
Db     966 GGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAACGCCTCCCTACTATGATG 1025

Qy     1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
            |  ||  ||||| || || |  ||| ||||||  || || || |||  |  |  |||
Db     1026 TGTCAGCCATGACCGTGCCAATTGCAGTGTGGAACGGTGGCCATGACATCCTGGCTGATC 1085

Qy     1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
            |  |||| ||  |||||||||  |  |  |  |||||||  | |||||||| | ||||
Db     1086 CCCAAGATGTCGCAATGCTGCTTCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTC 1145

Qy     1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
            |   |    |||| |||| ||||||||||   |||||||  |||||      | |||||||
Db     1146 TTCCCTACAATCACCTGGACTTCATCTGGGCGATGGATGCGCCTCAGGAGGTTTACAATG 1205

Qy     1205 aaatcatccatctgatg 1221
            | ||  |    |||||
Db     1206 AGATAGTTACCATGATG 1222


RESULT  10
```

```
AK009479
LOCUS       AK009479     1343 bp    mRNA            HTC       08-FEB-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310022K08, full insert sequence.
ACCESSION   AK009479
VERSION     AK009479.1  GI:12844303
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310022K08.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods Enzymol. 303, 19-44 (1999)
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Kikuchi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T.,
            Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T.,
            Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Okazaki,Y., Muramatsu,M., Inoue,Y. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1343)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 25.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES              Location/Qualifiers
     source           1. .1343
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /db_xref="MGD:MGI:1896992"
                      /db_xref="MGD:MGI:1914967"
                      /clone="2310022K08"
                      /sex="male"
                      /tissue_type="tongue"
                      /clone_lib="RIKEN full-length enriched mouse cDNA library"
                      /dev_stage="adult"
     CDS              50. .1237
                      /note="putative"
                      /codon_start=1
                      /protein_id="BAB26316.1"
                      /db_xref="GI:12844304"
                      /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                      SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
                      FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                      GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
                      FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV
                      YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                      AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                      AED"
BASE COUNT      394 a    289 c    282 g    378 t
ORIGIN
```

Query Match 36.1%; Score 458.6; DB 192; Length 1343;
Best Local Similarity 63.6%; Pred. No. 2.8e-128;
Matches 698; Conservative 0; Mismatches 399; Indels 0; Gaps 0;

```
Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
          |||| ||||||  |||||||  ||||| | || ||.   |    || |||||  ||||||
Db    129 ACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTACTGGGGATATCCAAGTGAGG 188

Qy    185 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 244
          ||||||||||.|  ||||||||||| || || ||   ||  | || ||||||   ||
Db    189 AATATGAAGTTGTTACTGAAGATGGCTACATTCTGGGGGTCTATAGAATTCCTTATGGGA 248

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
              |  || |||| |  ||   || |||||||  |  ||||||||||  |  |||
Db    249 AGAAAAATTCTGAGAATATCGGCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTG 308

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
          |   || |  |||||||||| | || |||| |||| |||||| |||||||| ||||
Db    309 CATCAGCCACAAACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAG 368

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
          ||||||| | ||| |||||| |||||||||  | || ||  | ||||| || ||| |
Db    369 ATGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCCGGAAAAATG 428

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
                ||   ||||  || ||| |||||||||||||| | |||||| |||||||  |
Db    429 TATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGGCTAAAT 488

Qy    485 ttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaaagatctatt 544
          |||||||| ||    ||| |||| ||| | |||||||| || || || ||||  | |
Db    489 ATGACCTTCCAGCCACCATAGACTTCATTGTACAGAAAACTGGACAAGAGAAGATACACT 548

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
          |||| ||  | || |||||||||| || || |||||||| ||||| ||||  || |  |
Db    549 ATGTTGGTCACTCTCAGGGCACCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTC 608

Qy    605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
          ||||| | |||||||| | || || ||| ||||| ||  | || ||||| ||| || ||
Db    609 TGGCTAAAAAAATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAG 668
```

```
Qy     665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
           |||| |||   |  ||  |||   |  | ||  |  |  |  ||||||  | |||||
Db     669 AAAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGATATTTGGTA 728

Qy     725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccttttgtggcc 784
           | |||   ||  ||   || | |   ||       |  |||| ||   |  | ||   |
Db     729 ACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTTGGTACGGAAGTGTGCTCAC 788

Qy     785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
           || | | || ||||   | || || ||     | ||  |  |  |||||||  |||  |
Db     789 GGGAGCTGCTAGATCTTCTCTGCAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGA 848

Qy     845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
           | ||  | ||  |||| ||     ||| |||   |  || | ||  || |||||||||
Db     849 AAAACTTAAATGTGAGTCGCTTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTA 908

Qy     905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
           |||  |  || | ||||||   |||   |  || |||||  | || |  || | | |||
Db     909 CTCAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGCCTATAACT 968

Qy     965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
           |||| |||       | ||  ||  | | | |||||||  ||| |||      ||   |
Db     969 GGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAACGCCTCCCTACTATGATG 1028

Qy    1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
           |  ||  ||||| || || |  ||| ||||||  || || || |||  ||  |  |||
Db    1029 TGTCAGCCATGACCGTGCCAATTGCAGTGTGGAACGGTGGCCATGACATCCTGGCTGATC 1088

Qy    1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
           |  |||| ||   |||||||||  |  |  |  ||||||  | |||||||| | ||||
Db    1089 CCCAAGATGTCGCAATGCTGCTTCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTC 1148

Qy    1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
           |   |    |||| |||| ||||||||||   |||||| |||||    | |||||||
Db    1149 TTCCCTACAATCACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATG 1208

Qy    1205 aaatcatccatctgatg 1221
           | ||  |    |||||
Db    1209 AGATAGTTACCATGATG 1225
```

```
RESULT  11
AK009413
LOCUS       AK009413     1345 bp    mRNA            HTC       08-FEB-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310020E22, full insert sequence.
ACCESSION   AK009413
VERSION     AK009413.1  GI:12844193
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310020E22.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods Enzymol. 303, 19-44 (1999)
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Kikuchi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T.,
            Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T.,
            Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Okazaki,Y., Muramatsu,M., Inoue,Y. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1345)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 25.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES             Location/Qualifiers
     source          1. .1345
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /db_xref="MGD:MGI:1895774"
                     /db_xref="MGD:MGI:1914967"
                     /clone="2310020E22"
                     /sex="male"
                     /tissue_type="tongue"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             52. .1239
                     /note="putative"
                     /codon_start=1
                     /protein_id="BAB26272.1"
                     /db_xref="GI:12844194"
                     /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                     SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
                     FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                     GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
                     FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV
                     YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                     AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                     AED"
BASE COUNT      395 a    288 c    283 g    379 t
ORIGIN
```

```
  Query Match           36.1%;  Score 458.6;  DB 192;  Length 1345;
  Best Local Similarity  63.6%;  Pred. No. 2.8e-128;
  Matches  698;  Conservative    0;  Mismatches  399;  Indels    0;  Gaps    0;

Qy     125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
           |||| ||||||  |||||||  |||||  | || ||     |    || |||||  ||||||
Db     131 ACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTACTGGGGATATCCAAGTGAGG 190

Qy     185 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 244
           ||||||||||  |  ||||||||||| || || ||    ||  | || ||||||   ||
Db     191 AATATGAAGTTGTTACTGAAGATGGCTACATTCTGGGGGTCTATAGAATTCCTTATGGGA 250

Qy     245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
           |  || |||| |  ||    || |||||||  |   ||||||||||  |  |||
Db     251 AGAAAAATTCTGAGAATATCGGCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTG 310

Qy     305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
           |  || |  |||||||| | || ||||| |||||| ||||||||  |||| |||| ||||
Db     311 CATCAGCCACAAACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAG 370

Qy     365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
           ||||||| | ||| |||||| ||||||||||  | || ||  | ||||| || ||| |
Db     371 ATGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCCGGAAAAATG 430

Qy     425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
           |   ||   ||||  || ||| |||||||||||||| | |||||| |||||||  |
Db     431 TATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGGCTAAAT 490

Qy     485 ttgaccttcctgcagtgataaactttatttttgcagaaaacgggccaggaaaagatctatt 544
           |||||||| ||    ||| |||| ||| | |||||||| || || || |||||  | |
Db     491 ATGACCTTCCAGCCACCATAGACTTCATTGTACAGAAAACTGGACAAGAGAAGATACACT 550

Qy     545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
           |||| ||  |  || |||||||||||| || || |||||||| |||||  |||| |  |
Db     551 ATGTTGGTCACTCTCAGGGCACCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTC 610

Qy     605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
           ||||| | |||||||| | || || ||| |||||  ||  | || |||||  ||| || ||
Db     611 TGGCTAAAAAAATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAG 670

Qy     665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
           |||| |||   |  ||  |||  |  | ||  |  |  |  ||||||   | |||||  |
Db     671 AAAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGATATTTGGTA 730

Qy     725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
           | |||   ||  ||   || | |   ||      |   |||| ||   |  | ||   |
Db     731 ACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTTGGTACGGAAGTGTGCTCAC 790

Qy     785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
           || | | || ||||   | || || ||     | ||  |  |  |||||||  |||  |
Db     791 GGGAGCTGCTAGATCTTCTCTGCAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGA 850

Qy     845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
           | ||  | ||  |||| ||      ||| |||   |  || | ||  || |||||||||
Db     851 AAAACTTAAATGTGAGTCGCTTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTA 910

Qy     905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
           |||  |  || | ||||||   |||   |  || |||||  | || |  || | | |||
Db     911 CTCAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGCCTATAACT 970

Qy     965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
           ||||  |||       |  ||  ||  | | | |||||||  |||  |||      ||   |
Db     971 GGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAACGCCTCCCTACTATGATG 1030

Qy    1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
           |  ||  ||||| || || |  ||| |||||||  || || || |||   ||  |  |||
Db    1031 TGTCAGCCATGACCGTGCCAATTGCAGTGTGGAACGGTGGCCATGACATCCTGGCTGATC 1090

Qy    1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
           |  |||| ||  |||||||||  |  |  | |||||||  | |||||||| | ||||
Db    1091 CCCAAGATGTCGCAATGCTGCTTCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTC 1150

Qy    1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
           |   |   |||| |||| ||||||||||   |||||| |||||     | |||||||
Db    1151 TTCCCTACAATCACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATG 1210

Qy    1205 aaatcatccatctgatg 1221
           | ||  |     |||||
Db    1211 AGATAGTTACCATGATG 1227
```

RESULT 12
AK009473
LOCUS AK009473 1345 bp mRNA HTC 08-FEB-2001
DEFINITION Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310022H17, full insert sequence.
ACCESSION AK009473
VERSION AK009473.1 GI:12844293
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library clone:2310022H17.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (sites)
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Methods Enzymol. 303, 19-44 (1999)
REFERENCE 2 (sites)
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
REFERENCE 3 (sites)
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Kikuchi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Okazaki,Y., Muramatsu,M., Inoue,Y. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
REFERENCE 4 (sites)
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5 (bases 1 to 1345)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Oxido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Gemomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for

further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 25.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI. Host: SOLR.

```
FEATURES             Location/Qualifiers
     source          1. .1345
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /db_xref="MGD:MGI:1898391"
                     /db_xref="MGD:MGI:1914967"
                     /clone="2310022H17"
                     /sex="male"
                     /tissue_type="tongue"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             52. .1239
                     /note="putative"
                     /codon_start=1
                     /protein_id="BAB26312.1"
                     /db_xref="GI:12844294"
                     /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                     SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
                     FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                     GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
                     FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV
                     YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                     AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                     AED"
BASE COUNT      395 a    288 c    283 g    379 t
ORIGIN
```

```
  Query Match           36.1%;  Score 458.6;  DB 192;  Length 1345;
  Best Local Similarity   63.6%;  Pred. No. 2.8e-128;
  Matches  698;  Conservative    0;  Mismatches  399;  Indels    0;  Gaps    0;

Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
          |||| ||||||  |||||||  |||||  | || ||      |    || |||||  ||||||
Db    131 ACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTACTGGGGATATCCAAGTGAGG 190

Qy    185 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 244
          ||||||||||  |  ||||||||||| || || ||    ||  | || ||||||   ||
Db    191 AATATGAAGTTGTTACTGAAGATGGCTACATTCTGGGGGTCTATAGAATTCCTTATGGGA 250

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
              |  || |||| |  ||    || |||||||  |   ||||||||||  |  |||
Db    251 AGAAAAATTCTGAGAATATCGGCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTG 310

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
          |   || |  |||||||||  | || |||||  |||||  |||||||  |||||||||  ||||
Db    311 CATCAGCCACAAACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAG 370

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
          |||||||  | |||  ||||||  ||||||||||   | ||  ||   | |||||  || |||  |
Db    371 ATGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCCGGAAAAATG 430

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
                ||   ||||  ||  |||  |||||||||||||  | ||||||  |||||||  |
Db    431 TATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGGCTAAAT 490

Qy    485 ttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 544
```

```
          |||||||||  ||    |||  ||||  |||  |  ||||||||  ||  ||  ||  |||||  |  |
Db    491 ATGACCTTCCAGCCACCATAGACTTCATTGTACAGAAAACTGGACAAGAGAAGATACACT 550

Qy    545 atgtgggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
          ||||  ||  |  ||  ||||||||||||  ||  ||  |||||||  |||||  ||||   ||  |  |
Db    551 ATGTTGGTCACTCTCAGGGCACCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTC 610

Qy    605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
          |||||  |  ||||||||  |  ||  ||  |||  |||||  ||   |  ||  |||||  |||  ||  ||
Db    611 TGGCTAAAAAAATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAG 670

Qy    665 aaaagcccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
          ||||  |||   |  ||  |||  |  |  ||  |  |  |  ||||||   |  ||||||  |
Db    671 AAAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGATATTTGGTA 730

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
          |  |||   ||  ||  ||  |  |  ||      |  ||||  ||   |  |  ||   |
Db    731 ACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTTGGTACGGAAGTGTGCTCAC 790

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
          ||  |  |  ||  ||||   |  ||  ||  ||      |  ||   |   |  |||||||  |||  |
Db    791 GGGAGCTGCTAGATCTTCTCTGCAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGA 850

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
          |  ||   |  ||  ||||  ||       |||  |||   |   ||  |  ||  ||  ||||||||||
Db    851 AAAACTTAAATGTGAGTCGCTTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTA 910

Qy    905 tgcaaaatattctacactggagccaggcagtgaatttctggtgaactccgggcatttgact 964
          |||  |  ||  |  ||||||   |||   |  ||  |||||  |  ||  |  ||  |  |||
Db    911 CTCAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGCCTATAACT 970

Qy    965 ggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
          ||||  |||        |  ||  ||  |  |  |  |||||||  |||  |||       ||   |
Db    971 GGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAACGCCTCCCTACTATGATG 1030

Qy    1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
           |   ||  ||||| ||  ||  |  |||  ||||||  ||  ||  ||  |||   ||  |  |||
Db    1031 TGTCAGCCATGACCGTGCCAATTGCAGTGTGGAACGGTGGCCATGACATCCTGGCTGATC 1090

Qy    1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
           |  ||||  ||   ||||||||  |  |  |  |||||||  |  |||||||| | ||||
Db    1091 CCCAAGATGTCGCAATGCTGCTTCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTC 1150

Qy    1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
           |  |   |||| |||| ||||||||||   |||||||  ||||     | |||||||
Db    1151 TTCCCTACAATCACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATG 1210

Qy    1205 aaatcatccatctgatg 1221
           |  ||  |     |||||
Db    1211 AGATAGTTACCATGATG 1227
```

```
RESULT  13
AK009359
LOCUS       AK009359      1346 bp    mRNA            HTC       08-FEB-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310015H15, full insert sequence.
ACCESSION   AK009359
VERSION     AK009359.1  GI:12844108
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310015H15.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods Enzymol. 303, 19-44 (1999)
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
```

```
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Kikuchi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T.,
            Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T.,
            Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Okazaki,Y., Muramatsu,M., Inoue,Y. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1346)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 25.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES              Location/Qualifiers
     source           1. .1346
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /db_xref="MGD:MGI:1895805"
                      /db_xref="MGD:MGI:1914967"
                      /clone="2310015H15"
                      /sex="male"
                      /tissue_type="tongue"
                      /clone_lib="RIKEN full-length enriched mouse cDNA library"
                      /dev_stage="adult"
     CDS              53. .1240
                      /note="putative"
                      /codon_start=1
                      /protein_id="BAB26240.1"
                      /db_xref="GI:12844109"
                      /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                      SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASAKNWITNLPNNSLA
                      FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                      GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKIHFIPK
                      FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV
                      YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                      AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                      AED"
BASE COUNT      396 a    289 c    282 g    379 t
ORIGIN
```

```
  Query Match             36.1%;  Score 458.6;  DB 192;  Length 1346;
  Best Local Similarity   63.6%;  Pred. No. 2.8e-128;
  Matches  698;  Conservative    0;  Mismatches  399;  Indels    0;  Gaps    0;

Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
          |||| ||||||  |||||||  |||||  | || ||      |    || |||||  ||||||
Db    132 ACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTACTGGGGATATCCAAGTGAGG 191

Qy    185 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 244
          ||||||||||  |  ||||||||||| || || ||    ||  | || ||||||   ||
Db    192 AATATGAAGTTGTTACTGAAGATGGCTACATTCTGGGGGTCTATAGAATTCCTTATGGGA 251

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
              |  || |||| |  ||    || |||||||  |  ||||||||| |  |||
Db    252 AGAAAAATTCTGAGAATATCGGCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTG 311

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
          |   || |  |||||||||  | || ||||| ||||| |||||| |||||||| ||||
Db    312 CATCAGCCAAAAACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAG 371

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
          ||||||| | ||| |||||| ||||||||||  | || ||  | ||||| || ||| |
Db    372 ATGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCCGGAAAAATG 431

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
                ||   ||||  || ||| |||||||||||||| | |||||| |||||||  |
Db    432 TATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGGCTAAAT 491

Qy    485 ttgaccttcctgcagtgataaacttattttgcagaaaacgggccaggaaaagatctatt 544
          |||||||| ||    ||| |||| ||| | |||||||| || || || |||||  | |
Db    492 ATGACCTTCCAGCCACCATAGACTTCATTGTACAGAAAACTGGACAAGAGAAGATACACT 551

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
          |||| ||  |  || |||||||||||| || || |||||||| ||||| ||||  || |  |
Db    552 ATGTTGGTCACTCTCAGGGCACCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTC 611

Qy    605 tgcctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
          ||||| | |||||||| | || || ||| ||||| ||  |  || ||||| ||| || ||
Db    612 TGGCTAAAAAAATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAG 671

Qy    665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
          |||| |||   |  ||  ||  ||  | ||  |  |  |  ||||||   | |||||  |
Db    672 AAAGTCCCTTTAAAAAGATTCACTTTATTCCTAAGTTTCTTCTCAAGGTGATATTTGGTA 731

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
          | |||   ||  ||   || |  ||       |   |||| ||   |  | ||    |
Db    732 ACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTTGGTACGGAAGTGTGCTCAC 791

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
          || | | || ||||   | || || ||     | ||  |  | |||||||  |||  |
Db    792 GGGAGCTGCTAGATCTTCTCTGCAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGA 851

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
          | ||  | ||  |||| ||       ||| |||   |  || | ||  || |||||||||
Db    852 AAAACTTAAATGTGAGTCGCTTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTA 911

Qy    905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
            || |  || | ||||||   |||   |  || |||||  | || |  || | | |||
Db    912 CTCAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGCCTATAACT 971
```

```
Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
          |||| |||       | ||  ||  | | | |||||||  ||| |||      ||    |
Db    972 GGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAACGCCTCCCTACTATGATG 1031

Qy   1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
          |   ||  ||||| || || |  ||| ||||||  || || || |||   ||  |  |||
Db   1032 TGTCAGCCATGACCGTGCCAATTGCAGTGTGGAACGGTGGCCATGACATCCTGGCTGATC 1091

Qy   1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
          |  |||| ||   |||||||||  |  |  |  |||||||  | |||||||| | ||||
Db   1092 CCCAAGATGTCGCAATGCTGCTTCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTC 1151

Qy   1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
           |   |    |||| |||| ||||||||||   |||||||  |||||     | |||||||
Db   1152 TTCCCTACAATCACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATG 1211

Qy   1205 aaatcatccatctgatg 1221
          | ||  |     |||||
Db   1212 AGATAGTTACCATGATG 1228


RESULT  14
AK009523
LOCUS       AK009523     1346 bp    mRNA            HTC       08-FEB-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310028I09, full insert sequence.
ACCESSION   AK009523
VERSION     AK009523.1  GI:12844369
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310028I09.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods Enzymol. 303, 19-44 (1999)
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Kikuchi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T.,
            Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T.,
            Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Okazaki,Y., Muramatsu,M., Inoue,Y. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1346)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 25.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES             Location/Qualifiers
     source          1. .1346
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /db_xref="MGD:MGI:1895759"
                     /db_xref="MGD:MGI:1914967"
                     /clone="2310028I09"
                     /sex="male"
                     /tissue_type="tongue"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             53. .1240
                     /note="putative"
                     /codon_start=1
                     /protein_id="BAB26338.1"
                     /db_xref="GI:12844370"
                     /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                     SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
                     FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                     GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
                     FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV
                     YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                     AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                     AED"
BASE COUNT      395 a    289 c    283 g    379 t
ORIGIN


  Query Match           36.1%;  Score 458.6;  DB 192;  Length 1346;
  Best Local Similarity   63.6%;  Pred. No. 2.8e-128;
  Matches  698;  Conservative    0;  Mismatches  399;  Indels    0;  Gaps    0;

Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
          |||| ||||||  ||||||| ||||| | || ||     |    || |||||  ||||||
Db    132 ACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTACTGGGGATATCCAAGTGAGG 191

Qy    185 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 244
          |||||||||| |  |||||||||||| || || ||    ||  | || ||||||   ||
Db    192 AATATGAAGTTGTTACTGAAGATGGCTACATTCTGGGGGTCTATAGAATTCCTTATGGGA 251

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
              |  || |||| | ||    || |||||||  |   ||||||||||  |  |||
Db    252 AGAAAAATTCTGAGAATATCGGCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTG 311

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
            |   || |  |||||||||| | || ||||| ||||| ||||||| ||||||||| ||||
```

```
Db    312 CATCAGCCACAAACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAG 371

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
          ||||||| | ||| |||||| |||||||||| | || ||  | ||||| || ||| |
Db    372 ATGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCCGGAAAAATG 431

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
                ||   ||||  || ||| |||||||||||||| | |||||| |||||||  |
Db    432 TATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGGCTAAAT 491

Qy    485 ttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 544
          |||||||| ||    ||| |||| ||| | |||||||| || || || ||||  | |
Db    492 ATGACCTTCCAGCCACCATAGACTTCATTGTACAGAAAACTGGACAAGAGAAGATACACT 551

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
          |||| ||  | || |||||||||| || || |||||||| ||||| |||| || | |
Db    552 ATGTTGGTCACTCTCAGGGCACCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTC 611

Qy    605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
          ||||| | |||||||| | || || ||| ||||| ||  | || ||||| ||| || ||
Db    612 TGGCTAAAAAAAATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAG 671

Qy    665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
          |||| |||   |  ||  |||  |  | ||  |  |  |  ||||||   | ||||| |
Db    672 AAAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGATATTTGGTA 731

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccttgtggcc 784
          | |||   ||  ||  || | |  ||     |   |||| ||   |  | ||    |
Db    732 ACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTTGGTACGGAAGTGTGCTCAC 791

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
            || | | || ||||   | || || ||     | ||  |  |  |||||||  |||  |
Db    792 GGGAGCTGCTAGATCTTCTCTGCAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGA 851

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
          | ||  | ||  |||| ||      ||| |||   |  || | ||  || |||||||||
Db    852 AAAACTTAAATGTGAGTCGCTTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTA 911

Qy    905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
            ||| |  || | ||||||   |||   |  || |||||  | || |  || | | |||
Db    912 CTCAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGCCTATAACT 971

Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
          |||| |||       | ||  ||  | | | |||||||  ||| |||      ||    |
Db    972 GGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAACGCCTCCCTACTATGATG 1031

Qy   1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
          |   ||  ||||| || || |  ||| |||||| || || || |||   ||  |  |||
Db   1032 TGTCAGCCATGACCGTGCCAATTGCAGTGTGGAACGGTGGCCATGACATCCTGGCTGATC 1091

Qy   1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
          |  |||| ||   |||||||||  |  |  |  |||||||  | |||||||| | ||||  /
Db   1092 CCCAAGATGTCGCAATGCTGCTTCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTC 1151

Qy   1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
            |   |    |||| |||| ||||||||||   |||||||  |||||     | |||||||
Db   1152 TTCCCTACAATCACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATG 1211

Qy   1205 aaatcatccatctgatg 1221
          | ||  |    |||||
Db   1212 AGATAGTTACCATGATG 1228


RESULT  15
AK009729
LOCUS       AK009729     1346 bp    mRNA            HTC       08-FEB-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310040L03, full insert sequence.
ACCESSION   AK009729
VERSION     AK009729.1  GI:12844702
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310040L03.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods Enzymol. 303, 19-44 (1999)
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Kikuchi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T.,
            Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T.,
            Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Okazaki,Y., Muramatsu,M., Inoue,Y. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1346)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 25.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES            Location/Qualifiers
     source         1. .1346
                    /organism="Mus musculus"
                    /strain="C57BL/6J"
                    /db_xref="taxon:10090"
```

```
                     /db_xref="MGD:MGI:1898390"
                     /db_xref="MGD:MGI:1914967"
                     /clone="2310040L03"
                     /sex="male"
                     /tissue_type="tongue"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             53..1240
                     /note="putative"
                     /codon_start=1
                     /protein_id="BAB26466.1"
                     /db_xref="GI:12844703"
                     /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                     SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
                     FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                     GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
                     FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV
                     YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                     AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                     AED"
BASE COUNT      395 a    288 c    283 g    380 t
ORIGIN

  Query Match             36.1%;  Score 458.6;  DB 192;  Length 1346;
  Best Local Similarity   63.6%;  Pred. No. 2.8e-128;
  Matches  698;  Conservative    0;  Mismatches  399;  Indels    0;  Gaps    0;

Qy        125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
Db        132 ACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTACTGGGGATATCCAAGTGAGG 191

Qy        185 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 244
Db        192 AATATGAAGTTGTTACTGAAGATGGCTACATTCTGGGGGTCTATAGAATTCCTTATGGGA 251

Qy        245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
Db        252 AGAAAAATTCTGAGAATATCGGCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTG 311

Qy        305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
Db        312 CATCAGCCACAAACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAG 371

Qy        365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
Db        372 ATGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCCGGAAAAATG 431

Qy        425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
Db        432 TATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGGCTAAAT 491

Qy        485 ttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 544
Db        492 ATGACCTTCCAGCCACCATAGACTTCATTGTACAGAAAACTGGACAAGAGAAGATACACT 551

Qy        545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
Db        552 ATGTTGGTCACTCTCAGGGCACCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTC 611

Qy        605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
Db        612 TGGCTAAAAAAATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAG 671

Qy        665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
Db        672 AAAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGATATTTGGTA 731

Qy        725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
Db        732 ACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTTGGTACGGAAGTGTGCTCAC 791

Qy        785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
Db        792 GGGAGCTGCTAGATCTTCTCTGCAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGA 851

Qy        845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
Db        852 AAAACTTAAATGTGAGTCGCTTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTA 911

Qy        905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
Db        912 CTCAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGCCTATAACT 971

Qy        965 ggggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
Db        972 GGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAACGCCTCCCTACTATGATG 1031

Qy       1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
Db       1032 TGTCAGCCATGACCGTGCCAATTGCAGTGTGGAACGGTGGCCATGACATCCTGGCTGATC 1091

Qy       1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
Db       1092 CCCAAGATGTCGCAATGCTGCTTCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTC 1151

Qy       1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
Db       1152 TTCCCTACAATCACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATG 1211

Qy       1205 aaatcatccatctgatg 1221
Db       1212 AGATAGTTACCATGATG 1228

Search completed: August 24, 2001, 21:23:06
Job time: 5573 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: August 24, 2001, 20:38:19 ; Search time 211.8 Seconds
(without alignments)
6059.636 Million cell updates/sec

Title: US-09-333-159-45
Perfect score: 2044
Sequence: 1 gtcgacccacgcgtccgggg.....aaaaaaaaaagggcggccgc 2044

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601:*
1: /SIDS8/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2044	100.0	2044	22	AAF45131	Human TANGO 294 cD
2	1269	62.1	1269	22	AAF45132	Human TANGO 294 OR
3	489.2	23.9	1378	14	AAQ42310	RGL precursor. Or
4	470	23.0	1367	7	AAN60685	Sequence encoding
5	470	23.0	1367	7	AAN60566	Sequence encoding
6	470	23.0	1367	17	AAT58916	Human gastric lipa
7	464.4	22.7	1336	6	AAN50385	Rat prelingual lip
8	444.2	21.7	1528	15	AAQ68388	Canine gastric lip
9	444.2	21.7	1528	17	AAT58915	Dog gastric lipase
10	444.2	21.7	1531	15	AAQ68389	Canine gastric lip
11	286	14.0	699	22	AAF28679	Human protein HP03
12	286	14.0	1308	22	AAF28689	Human protein HP03
13	257.8	12.6	617	21	AAA44349	Human secreted exp
14	193.2	9.5	572	21	AAA43744	Mouse secreted exp
c 15	127.6	6.2	936	22	AAF58252	Oligonucleotide D1
c 16	127.6	6.2	936	22	AAF58254	Oligonucleotide D1
c 17	127.6	6.2	936	22	AAF58257	Oligonucleotide D1
c 18	127.6	6.2	936	22	AAF58259	Oligonucleotide D2
c 19	127.6	6.2	936	22	AAF58262	Oligonucleotide D2
c 20	127.6	6.2	938	22	AAF58255	Oligonucleotide D1
21	122	6.0	936	22	AAF58252	Oligonucleotide D1
22	122	6.0	936	22	AAF58254	Oligonucleotide D1
23	122	6.0	936	22	AAF58257	Oligonucleotide D1
24	122	6.0	936	22	AAF58259	Oligonucleotide D2
25	122	6.0	936	22	AAF58262	Oligonucleotide D2
26	122	6.0	938	22	AAF58255	Oligonucleotide D1
27	100	4.9	1718	21	AAZ34958	Soybean acid triac
28	89.4	4.4	1483	21	AAZ34956	Rice acid triacylg
29	84.8	4.1	859	21	AAZ34953	Corn acid triacylg
30	72	3.5	1438	21	AAZ34959	Soybean acid triac
31	58.6	2.9	244	22	AAF58238	Oligonucleotide D1
c 32	58	2.8	244	22	AAF58238	Oligonucleotide D1
33	55.8	2.7	705	21	AAZ34952	Catalpa acid triac
34	54	2.6	549	21	AAC93827	Cat flea hindgut a
35	51	2.5	509	21	AAZ34954	Corn acid triacylg
36	48.2	2.4	178	19	AAV35431	Cell death regulat
37	46.4	2.3	43676	21	AAZ60606	Nucleotide sequenc
c 38	42	2.1	8041	17	AAT10105	Adherence conferri
39	40.4	2.0	1140	20	AAX04759	DNA encoding a coc
40	40.4	2.0	1857	21	AAC59142	Human secreted pro
41	40.4	2.0	10628	22	AAF57301	P. falciparum FCR3
42	40.2	2.0	2021	19	AAV17087	Homo sapiens tubul
43	40	2.0	1937	21	AAA47445	Macaque TANGO 232
44	40	2.0	2019	21	AAF18286	Lung cancer associ
45	39.4	1.9	3189	20	AAX20284	Borrelia burgdorfe

ALIGNMENTS

RESULT 1
AAF45131
ID AAF45131 standard; cDNA; 2044 BP.
XX
AC AAF45131;
XX
DT 30-MAR-2001 (first entry)
XX
DE Human TANGO 294 cDNA.
XX
KW Gene therapy; TANGO protein; INTERCEPT protein; neurological disorder;
KW central nervous system; focal brain disorder; bipolar affective disorder;
KW global-diffuse cerebral disorder; cerebrovascular; Alzheimer's disease;
KW senile dementia; Huntington's disease; amyotrophic lateral sclerosis;
KW Parkinson's; Gilles de la Tourette's syndrome; hypertension; sleep;
KW neuropsychiatric; psychoactive substance use; anxiety; ss.
XX
OS Homo sapiens.
XX
PN WO200077239-A2.
XX
PD 21-DEC-2000.
XX
PF 24-MAY-2000; 2000WO-US14858.
XX
PR 14-JUN-1999; 99US-0333159.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI McCarthy SA, Fraser CC, Sharp JD, Barnes TM;
XX
DR WPI; 2001-032313/04.
DR P-PSDB; AAB66065.
XX

PT TANGO and INTERCEPT nucleic acids, proteins, and antibodies, useful for
PT screening assays and diagnostic assays and for the treatment of
PT neurological diseases such as Alzheimer's, Parkinson's and Huntington's
PT disease -
XX
PS Claim 1; Fig 6; 359pp; English.
XX
CC The present invention relates to TANGO or INTERCEPT proteins and coding
CC sequences (see AAF45121-F45136 and AAF45138-F45139 and AAB66031-B66057,
CC AAB66064-B66083 and AAB66085). The TANGO/INTERCEPT proteins and coding
CC sequences are useful for the treatment of neurological disorders such as
CC central nervous system (CNS) disorders, CNS-related disorders, focal
CC brain disorders, global-diffuse cerebral disorders and other
CC neurological and cerebrovascular disorders. The CNS disorders include
CC Alzheimer's disease, senile dementia, Huntington's disease, amyotrophic
CC lateral sclerosis, Parkinson's, Gilles de la Tourette's syndrome,
CC autonomic function disorders such as hypertension and sleep disorders,
CC neuropsychiatric disorders, psychoactive substance use disorders,
CC anxiety, and bipolar affective disorder.
XX
SQ Sequence 2044 BP; 618 A; 401 C; 460 G; 565 T; 0 other;

Query Match 100.0%; Score 2044; DB 22; Length 2044;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2044; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 gtcgacccacgcgtccggggaattgcagcaggaaaatatgtgaagagtttttaaacccac 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 gtcgacccacgcgtccggggaattgcagcaggaaaatatgtgaagagtttttaaacccac 60

Qy     61 aaattcttcttactttagaattagttgttacattggcaggaaaaaataaatgcagatgtt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 aaattcttcttactttagaattagttgttacattggcaggaaaaaataaatgcagatgtt 120

Qy    121 ggaccatgttggaaaccttgtcaagacagtggattgtctcacacagaatggaaatgtggc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ggaccatgttggaaaccttgtcaagacagtggattgtctcacacagaatggaaatgtggc 180

Qy    181 ttctgattctggtggcgtatatgttccagagaaatgtgaattcagtacatatgccaacta 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ttctgattctggtggcgtatatgttccagagaaatgtgaattcagtacatatgccaacta 240

Qy    241 aagctgtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 aagctgtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatc 300

Qy    301 cctgtgaggaatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 cctgtgaggaatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattc 360

Qy    361 ctcgaggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatg 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ctcgaggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatg 420

Qy    421 gcctagttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttca 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 gcctagttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttca 480

Qy    481 ttctggcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ttctggcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctc 540

Qy    541 gaaaacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgaga 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 gaaaacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgaga 600

Qy    601 tggctaggtttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaa 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 tggctaggtttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaa 660

Qy    661 agatctattatgtcggctattcacagggcaccaccatgggctttattgcattttccacca 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    661 agatctattatgtcggctattcacagggcaccaccatgggctttattgcattttccacca 720

Qy    721 tgccagagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgtta 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 tgccagagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgtta 780

Qy    781 agcatgcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggat 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 agcatgcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggat 840

Qy    841 tgtttggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacc 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 tgtttggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacc 900

Qy    901 tttgtggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggat 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 tttgtggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggat 960

Qy    961 tcaacaccaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctg 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 tcaacaccaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctg 1020

Qy   1021 gaacatctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccggg 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 gaacatctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccggg 1080

Qy   1081 catttgactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaa 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 catttgactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaa 1140

Qy   1141 ggtacagagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggc 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 ggtacagagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggc 1200

Qy   1201 tttcaaatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccata 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 tttcaaatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccata 1260

Qy   1261 agaatattcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgta 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 agaatattcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgta 1320

Qy   1321 tgtacaatgaaatcatccatctgatgcagcaggaggagaccaacctttcccagggacggt 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 tgtacaatgaaatcatccatctgatgcagcaggaggagaccaacctttcccagggacggt 1380

Qy   1381 gtgaggccgtattgtgaagcatctgacactgacgatcttaggacaacctcctgagggatg 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 gtgaggccgtattgtgaagcatctgacactgacgatcttaggacaacctcctgagggatg 1440

Qy   1441 gggctaggacccatgaaggcagaattacggagagcagagacctagtatacatttttcaga 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 gggctaggacccatgaaggcagaattacggagagcagagacctagtatacatttttcaga 1500

Qy   1501 ttcctgcacttggcactaaatccgacacttacatttacatttttttttctgtaaattaaa 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 ttcctgcacttggcactaaatccgacacttacatttacatttttttttctgtaaattaaa 1560

Qy   1561 gtacttattaggtaaatagaggttttgtatgctattatatattctaccatcttgaagggt 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 gtacttattaggtaaatagaggttttgtatgctattatatattctaccatcttgaagggt 1620

Qy   1621 aggttttacctgatagccagaaaatatctagacattctctatatcattcaggtaaatctc 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 aggttttacctgatagccagaaaatatctagacattctctatatcattcaggtaaatctc 1680

Qy   1681 tttaaaacacctattgtttttttctataagccatatttttggagcactaaagtaaaatggc 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 tttaaaacacctattgtttttttctataagccatatttttggagcactaaagtaaaatggc 1740

Qy   1741 aaattgggacagatattgaggtctggagtctgtggattattgttgactttgacaaaataa 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 aaattgggacagatattgaggtctggagtctgtggattattgttgactttgacaaaataa 1800
```

```
Qy   1801 gctagacattttcaccttgttgccacagagacataacactacctcaggaagctgagctgc 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 gctagacattttcaccttgttgccacagagacataacactacctcaggaagctgagctgc 1860

Qy   1861 tttaaggacaacaacaacaaaatcagtgttacagtatggatgaaatctatgttaagcatt 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 tttaaggacaacaacaacaaaatcagtgttacagtatggatgaaatctatgttaagcatt 1920

Qy   1921 ctcagaataaggccaagttttatagttgcatctcagggaagaaaattttataggatgttt 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 ctcagaataaggccaagttttatagttgcatctcagggaagaaaattttataggatgttt 1980

Qy   1981 atgagttctccaataaatgcattctgcattacataaaaaaaaaaaaaaaaaaaaggcgg 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 atgagttctccaataaatgcattctgcattacataaaaaaaaaaaaaaaaaaaaggcgg 2040

Qy   2041 ccgc 2044
          ||||
Db   2041 ccgc 2044
```

RESULT 2
AAF45132
ID AAF45132 standard; cDNA; 1269 BP.
XX
AC AAF45132;
XX
DT 30-MAR-2001 (first entry)
XX
DE Human TANGO 294 ORF.
XX
KW Gene therapy; TANGO protein; INTERCEPT protein; neurological disorder;
KW central nervous system; focal brain disorder; bipolar affective disorder;
KW global-diffuse cerebral disorder; cerebrovascular; Alzheimer's disease;
KW senile dementia; Huntington's disease; amyotrophic lateral sclerosis;
KW Parkinson's; Gilles de la Tourette's syndrome; hypertension; sleep;
KW neuropsychiatric; psychoactive substance use; anxiety; ss.
XX
OS Homo sapiens.
XX
PN WO200077239-A2.
XX
PD 21-DEC-2000.
XX
PF 24-MAY-2000; 2000WO-US14858.
XX
PR 14-JUN-1999; 99US-0333159.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI McCarthy SA, Fraser CC, Sharp JD, Barnes TM;
XX
DR WPI; 2001-032313/04.
DR P-PSDB; AAB66065.
XX
PT TANGO and INTERCEPT nucleic acids, proteins, and antibodies, useful for
PT screening assays and diagnostic assays and for the treatment of
PT neurological diseases such as Alzheimer's, Parkinson's and Huntington's
PT disease -
XX
PS Claim 1; Fig 6; 359pp; English.
XX
CC The present invention relates to TANGO or INTERCEPT proteins and coding
CC sequences (see AAF45121-F45136 and AAF45138-F45139 and AAB66031-B66057,
CC AAB66064-B66083 and AAB66085). The TANGO/INTERCEPT proteins and coding
CC sequences are useful for the treatment of neurological disorders such as
CC central nervous system (CNS) disorders, CNS-related disorders, focal
CC brain disorders, global-diffuse cerebral disorders and other
CC neurological and cerebrovascular disorders. The CNS disorders include
CC Alzheimer's disease, senile dementia, Huntington's disease, amyotrophic
CC lateral sclerosis, Parkinson's, Gilles de la Tourette's syndrome,
CC autonomic function disorders such as hypertension and sleep disorders,
CC neuropsychiatric disorders, psychoactive substance use disorders,
CC anxiety, and bipolar affective disorder.
XX
SQ Sequence 1269 BP; 358 A; 268 C; 309 G; 334 T; 0 other;

```
  Query Match             62.1%;  Score 1269;  DB 22;  Length 1269;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1269;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    126 atgttggaaaccttgtcaagacagtggattgtctcacacagaatggaaatgtggcttctg 185
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 atgttggaaaccttgtcaagacagtggattgtctcacacagaatggaaatgtggcttctg 60

Qy    186 attctggtggcgtatatgttccagagaaatgtgaattcagtacatatgccaactaaagct 245
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 attctggtggcgtatatgttccagagaaatgtgaattcagtacatatgccaactaaagct 120

Qy    246 gtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgt 305
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 gtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgt 180

Qy    306 gaggaatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctcga 365
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 gaggaatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctcga 240

Qy    366 ggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggccta 425
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggccta 300

Qy    426 gttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctg 485
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 gttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctg 360

Qy    486 gcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaa 545
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 gcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaa 420

Qy    546 cacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatggct 605
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 cacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatggct 480

Qy    606 aggtttgaccttcctgcagtgataaactttatttttgcagaaaacgggccaggaaaagatc 665
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 aggtttgaccttcctgcagtgataaactttatttttgcagaaaacgggccaggaaaagatc 540

Qy    666 tattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgcca 725
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 tattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgcca 600

Qy    726 gagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcat 785
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 gagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcat 660

Qy    786 gcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgttt 845
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 gcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgttt 720

Qy    846 ggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccttgt 905
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 ggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccttgt 780

Qy    906 ggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaac 965
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 ggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaac 840

Qy    966 accaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaaca 1025
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 accaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaaca 900

Qy   1026 tctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcattt 1085
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 tctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcattt 960
```

```
Qy   1086 gactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtac 1145
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 gactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtac 1020

Qy   1146 agagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttca 1205
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 agagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttca 1080

Qy   1206 aatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaat 1265
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 aatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaat 1140

Qy   1266 attcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtac 1325
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 attcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtac 1200

Qy   1326 aatgaaatcatccatctgatgcagcaggaggagaccaacctttcccagggacggtgtgag 1385
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 aatgaaatcatccatctgatgcagcaggaggagaccaacctttcccagggacggtgtgag 1260

Qy   1386 gccgtattg 1394
          |||||||||
Db   1261 gccgtattg 1269
```

RESULT 3
AAQ42310
ID AAQ42310 standard; DNA; 1378 BP.
XX
AC AAQ42310;
XX
DT 20-SEP-1993 (first entry)
XX
DE RGL precursor.
XX
KW Rabbit gastric lipase; RGL; pRGLN2.1; fat; bioconversion;
KW hydrolysis; transesterification; ds.
XX
OS Oryctolagus cuniculus.
XX
FH Key Location/Qualifiers
FT misc_RNA 120..1247
FT /*tag= a
FT /note= "claim 1; page 10-11"
XX
PN EP542629-A.
XX
PD 19-MAY-1993.
XX
PF 12-NOV-1992; 92EP-0403055.
XX
PR 13-NOV-1991; 91FR-0013948.
XX
PA (LJOU) INST RECH JOUVEINAL.
XX
PI Benicourt C, Blanchard C, Junien J;
XX
DR WPI; 1993-161080/20.
DR P-PSDB; AAR37302.
XX
PT Rabbit gastric lipase, its precursor and their DNA - useful for
PT treating conditions linked to gastric lipase deficiency, such as
PT mucoviscidiosis and pancreatic exocrine insufficiency
XX
PS Claim 1; Fig 6; 31pp; French.
XX
CC mRNA was isolated from rabbit gastric mucosal cells, converted to
CC cDNA and fragments cloned in pUC18. Recombinant plasmids were used
CC to transform E.coli MM294 and cells screened for reactivity with
CC probes corresp. to known parts of the RGL sequence. One positive
CC clone contd. plasmid pJ0101 which contd. a 1.35 kb PstI sequence.
CC This fragment was isolated, subjected to PCR and the amplification
CC prod. ligated with the 2.2 kb BglII-SacI fragment of pRU276 (contg.
CC a synthetic Tac promoter and transcription stop signal). The
CC recombinant plasmid was designated pRGLN2.1.
XX
SQ Sequence 1378 BP; 397 A; 287 C; 276 G; 418 T; 0 other;

Query Match 23.9%; Score 489.2; DB 14; Length 1378;
Best Local Similarity 65.0%; Pred. No. 1.5e-115;
Matches 722; Conservative 0; Mismatches 388; Indels 0; Gaps 0;

```
Qy    250 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 309
          |||| ||||     |||||||||||| | || ||    |    || || ||  ||||
Db    136 accctgaagtgaatatgaatattagtcagatgatttcctactggggatacccaagtgaaa 195

Qy    310 aatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctcgaggcc 369
          |||||||||| |  |||||||||| |||||||||   ||||||| ||||||   ||
Db    196 aatatgaagttgtgactgaagatggttatatccttgaggttaacagaattccttatggga 255

Qy    370 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 429
              |  |   ||| | |||     || || || || ||  |||||||||| |  |||
Db    256 agaaaaattcagggaacagaggccagagacccgtcgtatttttgcagcatggtttgcttg 315

Qy    430 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 489
          |  ||    |||||||| |||||||||||||||| ||||||| ||| ||||||||||
Db    316 catcagcctcaaactggatctccaacctgcccaacaacagcctggccttttattctggcag 375

Qy    490 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 549
          ||||||||| ||  |||||| |||||||||||||| |||||  |||||||  ||| | ||
Db    376 atgctggttatggtgtgtggctggggaacagcagaggaaatacctggtccagaagaaact 435

Qy    550 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 609
              ||   ||||   | ||| ||||||||||| || | |||||| |||||||  |
Db    436 tatactattcaccagactccgttgaattctgggcttttagctttgatgaaatggctaaat 495

Qy    610 ttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaaagatctatt 669
          |||||||||  ||    ||  |||| ||| |  || |||| || ||||||||| |  | |
Db    496 atgaccttccagccacaattgacttcattgtaaaggaaactggacaggaaaagctgcact 555

Qy    670 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 729
          |||| ||  |||| ||||||||||||||| || |||||||| |||||  ||  |  |||
Db    556 atgttggtcattcccagggcaccaccattggttttattgccttttctacaaatcccaagc 615

Qy    730 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 789
          |||||  |  | |||||||| | || ||| |||||  ||  | |||||||| ||| || |||
Db    616 tggctgaaagaatcaaaaccttttatgcattagctccagtcgccactgtgaagtatacaa 675

Qy    790 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 849
          ||||||  |  | |||| ||  |||  | ||     |  ||| |||||     | |||||
Db    676 aaagccttgtaaacaaacttaggtttattcctccaaccatgttcaagattatatttggtg 735

Qy    850 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccttttgtggcc 909
          | |||  |||        || |     |||     |   ||||  | | |  | |||  ||
Db    736 acaaaatattctacccacacaatttctttgatcaatttcttgccactcaagtgtgttccc 795

Qy    910 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 969
            | ||  ||  ||   ||||| || |||  | | ||  |  |  |||| ||  ||| |
Db    796 gtgagacactgaatgtaatttgcagcaatgccttatttatcatttgtggctttgacagcg 855

Qy    970 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 1029
            ||  |||||||||| ||      ||| || |   | || | ||  || |||| || |
Db    856 caaacttgaacatgagtcgcttggatgtgtacgtgtcacataatccggcaggaacttcag 915

Qy   1030 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 1089
          | ||||| || || ||||||| |||||| || || |||||  |  | |  || ||| | |
Db    916 ttcaaaacatgctgcactggacccaggctgttaaatctggaaattttcaagctttaatt 975

Qy   1090 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1149
          |||| |||    |  | ||| | |   | | |||||||| || |||      |||| |
Db    976 ggggaagtccagctcagaatgtagtgcacttcaatcagcccacacctccctactacaatg 1035

Qy   1150 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1209
          | |  |  |||| || || |  ||| ||||||  || ||  | ||||||  |  |  | |
```

```
Db   1036 tgaccgccatgaatgtgccaattgcagtgtggagtggtggcaatgactggttggctgacc 1095

Qy   1210 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1269
          |  |||| ||  |  | |||||  |  |  |  | || ||||| ||||| ||| | ||||
Db   1096 cccaagatgttgacctttgcttccaaaactttctaatctcatttaccacaaggagattc 1155

Qy   1270 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1329
            |  ||    |||| ||||||| |||||||   || ||||||||||  . . | |||||||
Db   1156 ttccatataatcacttggattttatctgggcaatgaatgctcctcaagaagtttacaatg 1215

Qy   1330 aaatcatccatctgatgcagcaggaggaga 1359
          |||| ||   | |||||    | ||  | |
Db   1216 aaattatttctatgatggcaaaagataaaa 1245


RESULT   4
AAN60685
ID   AAN60685 standard; DNA; 1367 BP.
XX
AC   AAN60685;
XX
DT   08-JUN-1991  (first entry)
XX
DE   Sequence encoding pregastric lipase.
XX
KW   Precursor polypeptide; secretion vector; enzyme; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   sig_peptide     47..103
FT                   /*tag=  a
FT   mat_peptide     104..1243
FT                   /*tag=  b
XX
PN   WO8603778-A.
XX
PD   03-JUL-1986.
XX
PF   23-DEC-1985;    85WO-GB00599.
XX
PR   21-DEC-1984;    84GB-0032483.
PR   23-DEC-1985;    85WO-GB00599.
PR   01-JAN-1986;    86GB-0019568.
XX
PA   (BREW-) BREWING RES FOUND.
PA   (TUBB/) TUBB R S.
XX
PI   Tubb RS;
XX
DR   WPI; 1986-182910/28.
DR   P-PSDB; AAP60724.
XX
PT   New precursor polypeptide of defined sequence - and corresp. DNA
PT   used to transform hosts for prodn. of the polypeptide
XX
PS   Example; Fig 5; 60pp; English.
XX
CC   The inventors claim a DNA sequence encoding AAP60723 linked to a
CC   promoter upstream and a gene for a polypeptide downstream.
CC   Particular examples are the yeast emzyme AMG, the mammalian enzyme,
CC   gastric lipase and the mammalian lymphokine, interferon-alpha2.
XX
SQ   Sequence 1367 BP; 397 A; 284 C; 269 G; 417 T; 0 other;


  Query Match             23.0%;  Score 470;  DB 7;  Length 1367;
  Best Local Similarity   64.2%;  Pred. No. 1.2e-110;
  Matches  707;  Conservative    0;  Mismatches  395;  Indels    0;  Gaps    0;

Qy    244 ctgtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccct 303
          |||    ||| ||||     ||||| |||||| | || ||     ||   || || ||
Db    123 ctggaagccctgaagtgactatgaacattagtcagatgattacttattggggatacccaa 182

Qy    304 gtcaggaatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctc 363
            ||| |||||||||| |  |||||||||| ||||| |||   || || || ||||||
Db    183 atcaagaatatgaagttgtgactgaagatggttatattcttgaagtcaatagaattcctt 242

Qy    364 gacgcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcc 423
            ||        |  |   ||| |||||    || ||||| |||||  ||||||||||
Db    243 atcggaagaaaaattcagggaatacaggccagagacctgttgtgtttttgcagcatggtt 302

Qy    424 tacttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattc 483
          |  ||| |   || |  ||||||||||||||||||||| ||||| ||||| | ||||||||
Db    303 tgcttgcatcagccacaaactggatttccaacctgccgaacaacagccttgccttcattc 362

Qy    484 tggcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaa 543
          |||||||||||||| ||| |||||| |||| |||||||| |||||| |||| |  |||
Db    363 tggcagatgctggttatgatgtgtggctgggcaacagcagaggaaacacctgggccagaa 422

Qy    544 aacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatgg 603
            | ||  |       ||   |||   || ||| |||||||||||||| | |||||| ||||
Db    423 gaaacttgtactattcaccagattcagttgaattctgggctttcagctttgatgaaatgg 482

Qy    604 ctaggtttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaaaga 663
          |||  | |||||||||| ||    ||  |||| ||| |  |||||||| || ||| |  ||
Db    483 ctaaatatgaccttccagccacaatcgacttcattgtaaagaaaactggacagaagcagc 542

Qy    664 tctattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgc 723
          |  | ||||| ||| |||| |||||||||||||| || |||||||| ||||||||||  |
Db    543 tacactatgttggccattcccagggcaccaccattggttttattgccttttccaccaatc 602

Qy    724 cagagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagc 783
          |    |||||| | | |||||||| |  | |||| |||| ||  | |||||||| |||
Db    603 ccagcctggctaaaagaatcaaaaccttctatgctctagctcctgttgccactgtgaagt 662

Qy    784 atgcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgt 843
          || |||||||||      | |||| ||   ||  | ||  |      |  |||||    | |
Db    663 atacaaaaagccttataaacaaacttagatttgttcctcaatccctcttcaagtttatat 722

Qy    844 ttggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttt 903
          ||||  | |||  |||       || | |  |||     |  |||| || | |  | |
Db    723 ttggtgacaaaatattctacccacacaacttctttgatcaatttcttgctactgaagtgt 782

Qy    904 gtggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattca 963
          |  ||  | ||| ||  |||   |||| || |||  | | ||  |  |  |||||||
Db    783 gctcccgtgagatgctgaatctcctttgcagcaatgccttatttataatttgtggatttg 842

Qy    964 acaccaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaa 1023
          |||  || ||  | |||| ||| ||      ||| |||  . | || | ||  || ||||
Db    843 acagtaagaactttaacacgagtcgcttggatgtgtatctatcacataatccagcaggaa 902

Qy   1024 catctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcat 1083
          | ||||| ||||| ||  | || |||| |||||| || || ||||  || |||  || |
Db    903 cttctgttcaaaacatgttccattggacccaggctgttaagtctgggaaattccaagctt 962

Qy   1084 ttgactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggt 1143
            |||||||||  ||        | |||  |    | |   |||||  |    |||     |
Db    963 atgactggggaagcccagttcagaataggatgcactatgatcagtcccaacctccctact 1022

Qy   1144 acagagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggcttt 1203
          |||  || | ||  ||||  || || |  ||| |||||||  || ||  |||||  | |
Db   1023 acaatgtgacagccatgaatgtaccaattgcagtgtggaacggtggcaaggacctgttgg 1082

Qy   1204 caaatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataaga 1263
          |  | ||  |||| ||    |  ||||  |  |  |  |||| ||  || ||||| |||
Db   1083 ctgacccccaagatgttggccttttgcttccaaaactcccccaatcttatttaccacaagg 1142

Qy   1264 atattcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgt 1323
          | ||||||   |    |||| |||| || |||||||   |||||| |||||     | |
Db   1143 agattccttttttacaatcacttggactttatctgggcaatggatgcccctcaagaagttt 1202

Qy   1324 acaatgaaatcatccatctgat 1345
          ||||||| ||  |   | ||||
Db   1203 acaatgacattgtttctatgat 1224
```

```
RESULT   5
AAN60566
ID   AAN60566 standard; DNA; 1367 BP.
XX
AC   AAN60566;
XX
DT   22-AUG-1991  (first entry)
XX
DE   Sequence encoding human pregastric lipase.
XX
KW   Cystic fibrosis therapy; enzyme; lipase deficiency; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             47..103
FT                   /*tag= a
FT   mat_peptide     104..1243
FT                   /*tag= b
XX
PN   WO8601532-A.
XX
PD   13-MAR-1986.
XX
PF   15-AUG-1985;   85WO-GB00364.
XX
PR   21-AUG-1984;   84GB-0021210.
PR   15-AUG-1985;   85WO-GB00364.
PR   01-JAN-1986;   86GB-0008897.
XX
PA   (CELL-) CELLTECH LTD.
PA   (LOWE/) LOWE P A.
XX
PI   Lowe PA;
XX
DR   WPI; 1986-081634/12.
DR   P-PSDB; AAP60658.
XX
PT   New gastric lipase protein, esp. of human origin - for treating
PT   lipase deficiency, and DNA sequences coding for it
XX
PS   Disclosure; Fig 3; 39pp; English.
XX
CC   The inventors claim a pregastric lipase protein and a gene encoding
CC   it. Gastric lipase protein is useful for oral administration to
CC   treat lipase deficiency, e.g. cystic fibrosis or pancreatitis.
XX
SQ   Sequence 1367 BP; 397 A; 284 C; 269 G; 417 T; 0 other;

  Query Match             23.0%;  Score 470;  DB 7;  Length 1367;
  Best Local Similarity   64.2%;  Pred. No. 1.2e-110;
  Matches  707;  Conservative    0;  Mismatches  395;  Indels    0;  Gaps    0;

Qy    244 ctgtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccct 303
          |||     ||| ||||      ||||| |||||| | || ||     ||    || || ||
Db    123 ctggaagccctgaagtgactatgaacattagtcagatgattacttattggggatacccaa 182

Qy    304 gtgaggaatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctc 363
            ||| ||||||||||| |  |||||||||| ||||| |||     || || || ||||||
Db    183 atgaagaatatgaagttgtgactgaagatggttatattcttgaagtcaatagaattcctt 242

Qy    364 gaggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcc 423
            ||         |  |   ||| |||||   || ||||| |||||  ||||||||||
Db    243 atgggaagaaaaattcagggaatacaggccagagacctgttgtgtttttgcagcatggtt 302

Qy    424 tagttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattc 483
          |  ||| |    || |  ||||||||||||||||||| ||||| ||||| | ||||||||
Db    303 tgcttgcatcagccacaaactggatttccaacctgccgaacaacagccttgccttcattc 362

Qy    484 tggcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaa 543
          |||||||||||||| ||| |||||| |||| |||||||| |||||| ||||| |   |||
Db    363 tggcagatgctggttatgatgtgtggctgggcaacagcagaggaaacacctgggccagaa 422

Qy    544 aacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatgg 603
           | ||  |        ||   |||    || ||| |||||||||||||| | |||||| ||||
Db    423 gaaacttgtactattcaccagattcagttgaattctgggctttcagctttgatgaaatgg 482

Qy    604 ctacgtttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaaga 663
          |||  | |||||||||| ||     ||  ||||  |||  ||||||| || ||| |   ||
Db    483 ctaaatatgaccttccagccacaatcgacttcattgtaaagaaaactggacagaagcagc 542

Qy    664 tctattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgc 723
          |  | ||||| ||| |||| ||||||||||||| || |||||||| ||||||||||   |
Db    543 tacactatgttggccattcccagggcaccaccattggtttattgccttttccaccaatc 602

Qy    724 cagagctggctcagaaaatcaaaatgtatttgctttagcacccatagccactgttaagc 783
          |     |||||| | | ||||||||  |    | |||| |||| ||    | |||||||| |||
Db    603 ccagcctggctaaaagaatcaaaaccttctatgctctagctcctgttgccactgtgaagt 662

Qy    784 atgcaaaaagccccgggaccaaattttgttgctgccagatatgatgatcaagggattgt 843
          || |||||||||       | |||| ||    ||   | ||  |        | |||||     | |
Db    663 atacaaaaagccttataaacaaacttagatttgttcctcaatccctcttcaagtttatat 722

Qy    844 ttggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctt 903
          ||||   | |||  |||        | | |  ,|||         |  ||||  | |  |  |   |
Db    723 ttggtgacaaaatattctacccacacaacttctttgatcaatttcttgctactgaagtgt 782

Qy    904 gtggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattca 963
          |    ||  | |||  ||  |||     |||| || |||  | | ||  |   |  ||||||
Db    783 gctcccgtgagatgctgaatctcctttgcagcaatgccttatttataatttgtggatttg 842

Qy    964 acaccaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaa 1023
          |||  || ||  | |||| ||| ||          ||| |||     | || | ||  || ||||
Db    843 acagtaagaactttaacacgagtcgcttggatgtgtatctatcacataatccagcaggaa 902

Qy   1024 catctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcat 1083
          | ||||| ||||| ||   | || |||| ||||| || ||| ||||| || ||| || |
Db    903 cttctgttcaaaacatgttccattggacccaggctgttaagtctgggaaattccaagctt 962

Qy   1084 ttgactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggt 1143
          ||||||||| ||           | |||   | |    | |||||  |    |||   |
Db    963 atgactggggaagcccagttcagaataggatgcactatgatcagtcccaacctccctact 1022

Qy   1144 acagagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggcttt 1203
          |||  || | ||   ||||   || || |    ||| ||| ||||||  || ||   |||||  | |
Db   1023 acaatgtgacagccatgaatgtaccaattgcagtgtggaacggtggcaaggacctgttgg 1082

Qy   1204 caaatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataaga 1263
          |  | ||  |||| ||       |  ||||   |   |   |   |||| || || ||||| |||
Db   1083 ctgacccccaagatgttggccttttgcttccaaaactccccaatcttatttaccacaagg 1142

Qy   1264 atattcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgt 1323
           | ||||||        |||| |||| || ||||||| ||||||| |||||        | |
Db   1143 agattccttttacaatcacttggactttatctgggcaatggatgcccctcaagaagttt 1202

Qy   1324 acaatgaaatcatccatctgat 1345
          ||||||| ||   |    | ||||
Db   1203 acaatgacattgtttctatgat 1224


RESULT   6
AAT58916
ID   AAT58916 standard; DNA; 1367 BP.
XX
AC   AAT58916;
XX
DT   19-AUG-1997  (first entry)
XX
DE   Human gastric lipase coding sequence.
XX
KW   Duodenal; gastric; lipase; transgenic; plant; recombinant; extract; ss;
KW   food; absorption; fat; pancreatic; cystic fibrosis; exocrine; dairy;
KW   hydrolysis; trans-esterification; substrate; enzyme; biofuel; human.
XX
```

```
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             47..1243
FT                   /*tag= a
FT                   /product= gastric lipase
FT   sig_peptide     47..103
FT                   /*tag= b
FT   mat_peptide     104..1240
FT                   /*tag= c
XX
PN   WO9633277-A2.
XX
PD   24-OCT-1996.
XX
PF   19-APR-1996;   96WO-FR00606.
XX
PR   20-APR-1995;   95FR-0004754.
XX
PA   (BIOC-) BIOCEM SA.
PA   (LJOU ) INST RECH JOUVEINAL.
XX
PI   Baudino S,  Benicourt C,  Cudrey C,  Gruber V,  Lenee P;
PI   Merot B;
XX
DR   WPI; 1996-485783/48.
DR   P-PSDB; AAW09383.
XX
PT   Recombinant human or canine pre-duodenal lipase prodn. in transgenic
PT   plants - useful for facilitating absorption of fat, as bio-catalysts
PT   and for prodn. of bio-fuel
XX
PS   Claim 3; Fig 4; 130pp; French.
XX
CC   This is the nucleotide sequence encoding the human pre-duodenal (i.e.
CC   gastric) lipase enzyme.  The sequence can be used to generate transgenic
CC   plants producing recombinant lipase in an enzymatically active form.
CC   Alternatively bases encoding amino acids 20-23 or 20-73 (1-4 or 1-54 of
CC   the mature protein, respectively) can be deleted to form the derivatives
CC   designated delta-4 or delta-54 respectively.  Plants, or their extracts,
CC   expressing the lipases or the truncated derivatives, can be used:
CC   (a) as pharmaceuticals or food to facilitate absorption of fat, either
CC   in healthy subjects or in patients with inadequate levels of
CC   gastric/pancreatic lipase e.g. those with cystic fibrosis or exocrine
CC   pancreatic insufficient, in the elderly or in patients undergoing medical
CC   treatment which alters fat adsorption; (b) for performing industrial or
CC   agricultural reactions, e.g. in processing of fats or in the dairy
CC   industry, for hydrolysis or transesterification reactions, etc., where
CC   the plant material may provide both enzyme and substrate.  The transgenic
CC   plants can also be used for biofuel production.
XX
SQ   Sequence 1367 BP; 397 A; 284 C; 269 G; 417 T; 0 other;

  Query Match             23.0%;  Score 470;  DB 17;  Length 1367;
  Best Local Similarity   64.2%;  Pred. No. 1.2e-110;
  Matches  707;  Conservative    0;  Mismatches  395;  Indels    0;  Gaps    0;

Qy     244 ctgtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccct 303
           |||    ||| ||||     ||||| ||||||  | ||.||     ||   || || ||
Db     123 ctggaagccctgaagtgactatgaacattagtcagatgattacttattggggatacccaa 182

Qy     304 gtgaggaatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctc 363
           ||| ||||||||||| |  |||||||||| ||||| |||   || || || ||||||
Db     183 atgaagaatatgaagttgtgactgaagatggttatattcttgaagtcaatagaattcctt 242

Qy     364 gaggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcc 423
             ||        |  |   ||| |||||    || ||||| |||||  ||||||||||
Db     243 atgggaagaaaaattcagggaatacaggccagagacctgttgtgtttttgcagcatggtt 302

Qy     424 tagttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattc 483
           |  ||| |   || |  |||||||||||||||||| ||||| ||||| | ||||||||
Db     303 tgcttgcatcagccacaaactggatttccaacctgccgaacaacagccttgccttcattc 362

Qy     484 tggcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaa 543
           ||||||||||||||| ||| |||||| |||| |||||||| |||||| ||||| |  |||
Db     363 tggcagatgctggttatgatgtgtggctgggcaacagcagaggaaacacctgggccagaa 422

Qy     544 aacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatgg 603
             | ||  |      ||   |||   || ||| ||||||||||||| | |||||| ||||
Db     423 gaaacttgtactattcaccagattcagttgaattctgggctttcagctttgatgaaatgg 482

Qy     604 ctaggtttgaccttcctgcagtgataaactttatttttgcagaaaacgggccaggaaaaga 663
           |||  | |||||||| ||    ||  |||| ||| |  |||||||| || ||| |  ||
Db     483 ctaaatatgaccttccagccacaatcgacttcattgtaaagaaaactggacagaagcagc 542

Qy     664 tctattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgc 723
           |  | ||||| ||| |||| |||||||||||||| || |||||||| ||||||||||  |
Db     543 tacactatgttggccattcccagggcaccaccattggttttattgccttttccaccaatc 602

Qy     724 cagagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagc 783
           |    |||||| | | ||||||||  |  | |||| |||| ||  | ||||||||  |||
Db     603 ccagcctggctaaaagaatcaaaaccttctatgctctagctcctgttgccactgtgaagt 662

Qy     784 atgcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgt 843
           || ||||||||||     | |||| ||   ||  | ||. |      |  |||||    | |
Db     663 atacaaaaagccttataaacaaacttagatttgttcctcaatccctcttcaagtttatat 722

Qy     844 ttggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttt 903
           ||||  | |||  |||       || | |  |||      |  |||| || | |  | |
Db     723 ttggtgacaaaatattctacccacacaacttctttgatcaatttcttgctactgaagtgt 782

Qy     904 gtggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattca 963
           |   ||  | ||| ||  |||   |||| || |||  |.| ||  |  | ·|||||||
Db     783 gctcccgtgagatgctgaatctcctttgcagcaatgccttatttataatttgtggatttg 842

Qy     964 acaccaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaa 1023
           |||  || ||  | |||| |||.||      ||| |||    | || | ||  || ||||
Db     843 acagtaagaactttaacacgagtcgcttggatgtgtatctatcacataatccagcaggaa 902

Qy     1024 catctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcat 1083
            | ||||| |||| ||  | || |||| |||||| || || |||||  || |||  || |
Db     903 cttctgttcaaaacatgttccattggacccaggctgttaagtctgggaaattccaagctt 962

Qy     1084 ttgactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggt 1143
            |||||||| ||       | |||  |  | |  |||| |    |||      |
Db     963 atgactggggaagcccagttcagaataggatgcactatgatcagtcccaacctccctact 1022

Qy     1144 acagagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggcttt 1203
            |||  || | ||  |||| || || |  ||| |||||||  || ||  |||||  | |
Db     1023 acaatgtgacagccatgaatgtaccaattgcagtgtggaacggtggcaaggacctgttgg 1082

Qy     1204 caaatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataaga 1263
            |  ||  |||| ||     |  ||||  |  |  |  |||| || || ||||| |||
Db     1083 ctgacccccaagatgttggccttttgcttccaaaactccccaatcttatttaccacaagg 1142

Qy     1264 atattcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgt 1323
            | ||||||   |    |||| |||| || |||||||  |||||| |||||     | |
Db     1143 agattccttttacaatcacttggactttatctgggcaatggatgcccctcaagaagttt 1202

Qy     1324 acaatgaaatcatccatctgat 1345
            |||||||  ||  |   | ||||
Db     1203 acaatgacattgtttctatgat 1224


RESULT   7
AAN50385
ID   AAN50385 standard; DNA; 1336 BP.
XX
AC   AAN50385;
XX
DT   17-JAN-1992  (first entry)
XX
DE   Rat prelingual lipase gene.
XX
KW   Prelingual lipase; enzyme; EC-3.1.1.3; ss.
```

```
XX
OS   Rattus rattus.
XX
FH   Key             Location/Qualifiers
FT   CDS             7..1194
FT                   /*tag=  a
XX
PN   GB2142337-A.
XX
PD   16-JAN-1985.
XX
PF   29-JUN-1984;   84GB-0016581.
XX
PR   01-JUL-1983;   83GB-0017989.
PR   05-SEP-1983;   83GB-0023759.
XX
PA   (CELL-) CELLTECH LTD.
XX
PI   Carey NH,  Williamson R;
XX
DR   WPI; 1985-014450/03.
DR   P-PSDB; AAP50322.
XX
PT   New lingual lipase protein for treatment of lipase deficiency - also
PT   new pre-lingual lipase protein and related products
XX
PS   Disclosure; Fig 5; 15pp; English.
XX
CC   This sequence may be expressed in a transformant host organism using
CC   a vector plasmid in order to produce a prelingual lipase protein.
CC   The expressed protein may be used for the treatment of lipase
CC   deficiency.
XX
SQ   Sequence 1336 BP; 409 A; 285 C; 263 G; 379 T; 0 other;

  Query Match             22.7%;  Score 464.4;  DB 6;  Length 1336;
  Best Local Similarity   63.4%;  Pred. No. 3.2e-109;
  Matches  711;  Conservative    0;  Mismatches  411;  Indels    0;  Gaps    0;

Qy     244 ctgtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccct 303
           ||||  |||| ||||||  ||||||||||| | || ||     |    || ||||| |
Db      80 ctggaaaccctgaagcaaatatgaatattagtcagatgataacttactggggatatccat 139

Qy     304 gtgaggaatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctc 363
           || | ||||||||||| |  |||||||||||| || || ||    || |||| |||||||
Db     140 gtcaagaatatgaagttgttactgaagatggctacattctgggggtctacagaattcctc 199

Qy     364 gaggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcc 423
             ||       |  || | || | |||    || ||||||||||   ||||||||||
Db     200 atgggaagaataattctgaaaatataggcaagagacctgtggtgtatttgcagcatggtt 259

Qy     424 tagttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattc 483
           |  | | |   || |  |||||||||| | || || || ||||| ||||| | ||| || | |
Db     260 tgatcgcatcagccacaaactggattgcaaatctaccaaacaacagcctcgcctttatgc 319

Qy     484 tggcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaa 543
           | ||||| ||||| | ||| |||||| |||||||||  | |||||  | |||||  | |
Db     320 tagcagacgctggctatgatgtgtggctggggaacagtcgaggaaatacatggtccagga 379

Qy     544 aacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatgg 603
           || |         |||   ||||  || ||| ||||||||||||||| | |||||| ||||
Db     380 aaaatgtatactactcaccagactcagttgaattctgggctttcagctttgatgaaatgg 439

Qy     604 ctaggtttgaccttcctgcagtgataaactttatttgcagaaaacgggccaggaaaaga 663
           |||  | |||||||| ||    |||||||| ||| | |||||||| || || || ||||
Db     440 ctaaatatgaccttcccgccacaataaacttcattgtacagaaaactggacaagagaaga 499

Qy     664 tctattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgc 723
           |  | ||||| ||  |||| ||| ||||||| || || || ||||| ||||| ||||  |
Db     500 tacactatgttggtcattctcagcgcaccactattggtttcattgccttttctaccaatc 559

Qy     724 cagagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagc 783
```

```
           |    |||||  | |||||||| | || || ||| ||||| ||  | || || || |||
Db     560 ctacactggccaaaaaaatcaagacgttttatgcattagctccagttgctaccgtgaagt 619

Qy     784 atgcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgt 843
           || || |||| ||| ||  ||  |||  ||  | ||    |  |  |||||    |||
Db     620 atacacaaagtcccttgaaaaagatttcatttattcctacatttcttttcaagcttatgt 679

Qy     844 ttggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttt 903
           ||||||| |||   || |||   || |||   |||          |||| ||   |  | |
Db     680 ttggcaagaaaatgttcctgccccacacctactttgatgactttcttggtaccgaagtgt 739

Qy     904 gtggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattca 963
           |    | || | |||| ||||   | || || || ||  |  || |  |  |||||||
Db     740 gctcacgggaggttctagatcttctctgcagcaacactttattcatcttctgtggatttg 799

Qy     964 acaccaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaa 1023
           |||  || ||  | ||  |||| ||       ||| |||   |  || | ||  || || |
Db     800 acaagaaaaacttaaatgtgagtcgttttgatgtgtatctagggcataatccagcaggga 859

Qy    1024 catctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcat 1083
           |||.||| ||| |  |||| ||||||   |||   || |  |||||  || | |  || |
Db     860 catctgttcaagactttctccactgggcacagcttgttagatctgggaaatttcaagcct 919

Qy    1084 ttgactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggt 1143
           |  |||||||  ||     || | ||  ||  | | | ||| |||  ||| |||      |
Db     920 tcaactggggaagcccatcccagaacatgttacactacaaccagaaaacgcctcctgaat 979

Qy    1144 acagagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggcttt 1203
           |    ||   ||  ||||| || ||    ||| ||||||  || ||  | |||   ||
Db     980 atgatgtgtcagccatgactgtgccagttgcagtgtggaacggtggcaatgacatcctgg 1039

Qy    1204 caaatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataaga 1263
           |  ||||  |||| ||    ||||||||  |  |  |  |||||||| | | |||||||
Db    1040 ctgatccccaagatgtcgccatgctgcttcccaaactctccaacctcctgttccataagg 1099

Qy    1264 atattcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgt 1323
           | |||| ||  |    |||| |||| ||||||||||   ||||||| |||||     | |
Db    1100 agattcttgcctacaatcacctggacttcatctgggcaatggatgcccctcaagaggttt 1159

Qy    1324 acaatgaaatcatccatctgatgcagcaggaggagaccaacc 1365
           ||||||| || ||     |||||    | ||  | |  || |
Db    1160 acaatgagatgatttccatgatggcagaagactaaaagaatc 1201


RESULT   8
AAQ68388
ID   AAQ68388 standard; DNA; 1528 BP.
XX
AC   AAQ68388;
XX
DT   20-FEB-1995  (first entry)
XX
DE   Canine gastric lipase.
XX
KW   Canine gastric lipase; CGL; dog; fat; mucoviscidosis; enzyme;
KW   bioconversion; exocrine pancreatic insufficiency; ss.
XX
OS   Canis familiaris.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1140
FT                   /*tag=  a
FT   misc_RNA        1..1137
FT                   /*tag=  b
FT                   /note= "claim 1"
XX
PN   WO9413816-A.
XX
PD   23-JUN-1994.
XX
PF   16-DEC-1993;   93WO-FR01260.
XX
PR   16-DEC-1992;   92FR-0015201.
```

```
XX
PA   (LJOU ) INST RECH JOUVEINAL.
XX
PI   Benicourt C,  Blanchard C,  Junien J;
XX
DR   WPI; 1994-217890/26.
DR   P-PSDB; AAR56870.
XX
PT   Recombinant canine gastric lipase and nucleic acid encoding it -
PT   are used for improving absorption of ingested fat, treating
PT   mucoviscidosis etc. and in enzymatic bio-conversions
XX
PS   Claim 1; Fig 8; 52pp; French.
XX
CC   The sequence given below is the sequence of figure 8, altered
CC   according to the amendments described on page 2 of the appended
CC   letter.
CC   CGL is used to improve absorption of ingested fat, in healthy and
CC   sick patients (e.g. having altered levels of gastric lipase); to
CC   treat conditions associated with insufficiency (or lack) of lipases,
CC   esp. mucoviscidosis or exocrine pancreatic insufficiency and partic.
CC   where immobilised, for bioconversions, e.g. hydrolysis or
CC   transesterification (other mammalian gastric lipases, or derivs.,
CC   can be used in this application).
XX
SQ   Sequence 1528 BP; 394 A; 386 C; 328 G; 420 T; 0 other;


  Query Match          21.7%;  Score 444.2;  DB 15;  Length 1528;
  Best Local Similarity   62.8%;  Pred. No. 5.1e-104;
  Matches  689;  Conservative    0;  Mismatches  408;  Indels    0;  Gaps    0;

Qy    250 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 309
          |||| ||||    |||||||||| ||| | || |||    |    || || ||   |||||
Db     26 accctgaagtgaccatgaatataagtcagatgatcacctactggggatacccagctgagg 85

Qy    310 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 369
          |||||||||| |  || ||||| || |||||||||    |  |||| ||||||   ||
Db     86 aatatgaagttgtgaccgaagacggttatatccttgggatcgacagaattccttatggga 145

Qy    370 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 429
             |  |  |||| | |||    || ||||| |  ||  |||| || ||  |  | |
Db    146 ggaaaaattcagagaatataggccggagacctgttgcatttttgcaacacggtttgctcg 205

Qy    430 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 489
          |   || |  |||||||| |||||||||||||||||| ||||||| |||||| ||||| |
Db    206 catcagccacaaactggatctccaacctgcccaacaacagcctggccttcatcctggccg 265

Qy    490 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 549
          | || || |  |||||||||| |||||||||||||||||| ||| ||||| |  | |   |
Db    266 acgccgggtacgacgtgtggctggggaacagcaggggcaacacctgggccaggaggaatc 325

Qy    550 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 609
          |     |||    |||   |  || |||||||||||||| | ||| ||||||||||  |
Db    326 tgtactactcgcccgactccgtcgaattctgggctttcagctttgacgagatggctaaat 385

Qy    610 ttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaaagatctatt 669
          |||||||||  ||    ||  |||| || ||| |||||||||| ||||| ||| |  | |
Db    386 atgaccttcccgccaccattgacttcatcttgaagaaaacgggacaggacaagctacact 445

Qy    670 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 729
          | || ||| |||| ||||||||||||||| || || || || ||||||||||  ||  |||
Db    446 acgttggccattcccagggcaccaccattggtttcatcgccttttccaccaatcccaagc 505

Qy    730 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 789
          ||||  |    |||||||  |  | ||| ||||| ||| | ||||| || ||| |  |
Db    506 tggcgaaacggatcaaaaccttctatgcattagctcccgttgccaccgtgaagtacaccg 565

Qy    790 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 849
          ||| ||     | |||| |  || |  | ||     |  |  |||||    | ||||| |
Db    566 aaaccctgttaaacaaactcatgctcgtcccttcgttcctcttcaagcttatatttggaa 625

Qy    850 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 909
```

```
          | |||  |||       ||   |  |||     |  || |  |   |  | ||   ||
Db    626 acaaaatattctacccacaccacttctttgatcaatttctcgccaccgaggtatgctccc 685

Qy    910 agctgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 969
            | ||   | ||||   | || || ||   | ||||  |  |  |||||||  |||| |
Db    686 gccagacggtggatctcctctgcagcaacgccctgtttatcatttgtggatttgacacta 745

Qy    970 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 1029
            ||  |||||||||| ||      ||| |||  . | || | ||  || |||||||| |
Db    746 tgaacttgaacatgagtcgcttggatgtgtatctgtcacataatccagcaggaacatcgg 805

Qy   1030 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 1089
          | || ||  | || ||||||  ||||||  || || |||||  |  |||  || |||||||
Db    806 ttcagaacgtgctccactggtcccaggctgttaagtctgggaagttccaagcttttgact 865

Qy   1090 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1149
          ||||  ||         | ||  ||    | |   |||||   |  |||      ||||
Db    866 ggggaagcccagttcagaacatgatgcactatcatcagagcatgcctccctactacaacc 925

Qy   1150 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1209
          | | ||| |||   || || |  ||| ||||||  || ||  |  |||| |||  |  | |
Db    926 tgacagacatgcatgtgccaatcgcagtgtggaacggtggcaacgacttgctggccgacc 985

Qy   1210 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1269
          |  | || ||  |  |  |||| ||  || |  |||| ||||| ||||| | ||| ||||
Db    986 ctcacgatgttgacctttgctttccaagctcccaatctcatttaccacaggaagattc 1045

Qy   1270 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1329
          ||   |    |||| |||| || |||||||  |||||| |||||     | |||||||
Db   1046 ctccttacaatcacttggactttatctgggccatggatgcccctcaagcggtttacaatg 1105

Qy   1330 aaatcatccatctgatg 1346
          ||||  |     |||||
Db   1106 aaattgtttccatgatg 1122


RESULT   9
AAT58915
ID   AAT58915 standard; cDNA to mRNA; 1528 BP.
XX
AC   AAT58915;
XX
DT   19-AUG-1997  (first entry)
XX
DE   Dog gastric lipase coding sequence.
XX
KW   Dog; duodenal; gastric; lipase; transgenic; plant; recombinant; extract;
KW   food; absorption; fat; pancreatic; cystic fibrosis; exocrine; dairy;
KW   hydrolysis; trans-esterification; substrate; enzyme; biofuel; ss.
XX
OS   Canis familiaris.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1140
FT                   /*tag= a
FT                   /product= gastric lipase
XX
PN   WO9633277-A2.
XX
PD   24-OCT-1996.
XX
PF   19-APR-1996;   96WO-FR00606.
XX
PR   20-APR-1995;   95FR-0004754.
XX
PA   (BIOC-) BIOCEM SA.
PA   (LJOU ) INST RECH JOUVEINAL.
XX
PI   Baudino S,  Benicourt C,  Cudrey C,  Gruber V,  Lenee P;
PI   Merot B;
XX
DR   WPI; 1996-485783/48.
DR   P-PSDB; AAW09382.
```

```
XX
PT   Recombinant human or canine pre-duodenal lipase prodn. in transgenic
PT   plants - useful for facilitating absorption of fat, as bio-catalysts
PT   and for prodn. of bio-fuel
XX
PS   Claim 2; Fig 1; 130pp; French.
XX
CC   This is the nucleotide sequence encoding the dog pre-duodenal (i.e.
CC   gastric) lipase enzyme.  The sequence can be used to generate transgenic
CC   plants producing recombinant lipase in an enzymatically active form.
CC   Alternatively the first 12 bases of this sequence (encoding amino acids
CC   1-4 of the protein encoded by this sequence) or bases 1-162 of this
CC   sequence (encoding amino acids 1-54) can be deleted to form the
CC   derivatives designated delta-4 or delta-54 respectively.  Plants, or
CC   their extracts, expressing the lipases or the truncated derivatives, can
CC   be used: (a) as pharmaceuticals or food to facilitate absorption of fat,
CC   either in healthy subjects or in patients with inadequate levels of
CC   gastric/pancreatic lipase e.g. those with cystic fibrosis or exocrine
CC   pancreatic insufficient, in the elderly or in patients undergoing medical
CC   treatment which alters fat adsorption; (b) for performing industrial or
CC   agricultural reactions, e.g. in processing of fats or in the dairy
CC   industry, for hydrolysis or transesterification reactions, etc., where
CC   the plant material may provide both enzyme and substrate.  The transgenic
CC   plants can also be used for biofuel production.
XX
SQ   Sequence 1528 BP; 392 A; 387 C; 329 G; 420 T; 0 other;


  Query Match             21.7%;  Score 444.2;  DB 17;  Length 1528;
  Best Local Similarity   62.8%;  Pred. No. 5.1e-104;
  Matches  689;  Conservative    0;  Mismatches  408;  Indels    0;  Gaps    0;

Qy    250 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 309
          |||| ||||    |||||||||| ||| | || |||    |    || || ||   |||||
Db     26 accctgaagtgaccatgaatataagtcagatgatcacctactggggatacccagctgagg 85

Qy    310 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 369
          |||||||||| |  || ||||| || |||||||||    | ||||| ||||||   ||
Db     86 aatatgaagttgtgaccgaagacggttatatccttgggatcgacagaattccttatggga 145

Qy    370 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 429
             |  | |||| | |||    || ||||| |  ||  |||| || ||  |  | |
Db    146 ggaaaaattcagagaatataggccggagacctgttgcatttttgcaacacggtttgctcg 205

Qy    430 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 489
          |   || |  |||||||| ||||||||||||||||| ||||||| |||||| ||||| |
Db    206 catcagccacaaactggatctccaacctgcccaacaacagcctggccttcatcctggccg 265

Qy    490 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 549
          | || || |  |||||||| |||||||||||||||| ||| ||||| |  | |   |
Db    266 acgccgggtacgacgtgtggctggggaacagcaggggcaacacctgggccaggaggaatc 325

Qy    550 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 609
          |     |||    |||   |  || |||||||||||||| | ||| ||||||||||  |
Db    326 tgtactactcgcccgactccgtcgaattctgggctttcagctttgacgagatggctaaat 385

Qy    610 ttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 669
          |||||||||| ||    ||  |||| || ||| |||||||||| ||||| ||| |  | |
Db    386 atgaccttcccgccaccattgacttcatcttgaagaaaacgggacaggacaagctacact 445

Qy    670 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 729
          | || ||| |||| |||||||||||||| || || || || ||||||||||  ||  |||
Db    446 acgttggccattcccagggcaccaccattggtttcatcgccttttccaccaatcccaagc 505

Qy    730 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 789
          ||||  |    |||||||  |  | ||| ||||| ||| | ||||| || ||| |  |
Db    506 tggcgaaacggatcaaaaccttctatgcattagctcccgttgccaccgtgaagtacaccg 565

Qy    790 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 849
          ||| ||    | |||| |  || |  | ||    |  | |||||   | |||||  |
Db    566 aaaccctgttaaacaaactcatgctcgtcccttcgttcctcttcaagcttatatttggaa 625

Qy    850 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccttttgtggcc 909
```

```
                |  |||  |||       ||   |   |||      |   || |  |   |  | ||   ||
Db    626 acaaaatattctacccacaccacttctttgatcaatttctcgccaccgaggtatgctccc 685

Qy    910 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 969
            | ||   | ||||   | || || ||   | ||||  |  |  |||||||  |||| |
Db    686 gcgagacggtggatctcctctgcagcaacgccctgtttatcatttgtggatttgacacta 745

Qy    970 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 1029
            ||  |||||||||| ||       ||| |||    | || |  ||  || |||||||| |
Db    746 tgaacttgaacatgagtcgcttggatgtgtatctgtcacataatccagcaggaacatcgg 805

Qy   1030 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 1089
          | |  ||  | || ||||||  ||||||  || || |||||  |  |||  || |||||||
Db    806 ttcagaacgtgctccactggtcccaggctgttaagtctgggaagttccaagcttttgact 865

Qy   1090 ggggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1149
          ||||  ||        | ||  ||    | |   |||||    |  |||      ||||
Db    866 ggggaagcccagttcagaacatgatgcactatcatcagagcatgcctccctactacaacc 925

Qy   1150 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1209
          | | ||| |||   || || |  ||| ||||||  || ||  | |||| |||  |  | |
Db    926 tgacagacatgcatgtgccaatcgcagtgtggaacggtggcaacgacttgctggccgacc 985

Qy   1210 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1269
          |  | || ||  |  |  |||| ||  || |  |||| ||||| ||||| | ||| ||||
Db    986 ctcacgatgttgacctttttgctttccaagctccccaatctcatttaccacaggaagattc 1045

Qy   1270 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1329
            ||   |    |||| |||| || |||||||   |||||||  |||||     | |||||||
Db    1046 ctccttacaatcacttggactttatctgggccatggatgcccctcaagcggtttacaatg 1105

Qy   1330 aaatcatccatctgatg 1346
          ||||  |     |||||
Db    1106 aaattgtttccatgatg 1122


RESULT  10
AAQ68389
ID   AAQ68389 standard; DNA; 1531 BP.
XX
AC   AAQ68389;
XX
DT   20-FEB-1995  (first entry)
XX
DE   Canine gastric lipase (Met-CGL).
XX
KW   Canine gastric lipase; CGL; dog; fat; mucoviscidosis; enzyme;
KW   bioconversion; exocrine pancreatic insufficiency; ss.
XX
OS   Canis familiaris.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1143
FT                   /*tag=  a
XX
PN   WO9413816-A.
XX
PD   23-JUN-1994.
XX
PF   16-DEC-1993;   93WO-FR01260.
XX
PR   16-DEC-1992;   92FR-0015201.
XX
PA   (LJOU ) INST RECH JOUVEINAL.
XX
PI   Benicourt C,  Blanchard C,  Junien J;
XX
DR   WPI; 1994-217890/26.
DR   P-PSDB; AAR56871.
XX
PT   Recombinant canine gastric lipase and nucleic acid encoding it -
PT   are used for improving absorption of ingested fat, treating
PT   mucoviscidosis etc. and in enzymatic bio-conversions
```

```
XX
PS   Claim 2; Fig 8; 52pp; French.
XX
CC   The sequence given below is the sequence of figure 8, altered
CC   according to the amendments described on page 2 of the appended
CC   letter.
CC   CGL is used to improve absorption of ingested fat, in healthy and
CC   sick patients (e.g. having altered levels of gastric lipase); to
CC   treat conditions associated with insufficiency (or lack) of lipases,
CC   esp. mucoviscidosis or exocrine pancreatic insufficiency and partic.
CC   where immobilised, for bioconversions, e.g. hydrolysis or
CC   transesterification (other mammalian gastric lipases, or derivs.,
CC   can be used in this application).
XX
SQ   Sequence 1531 BP; 395 A; 386 C; 329 G; 421 T; 0 other;


  Query Match             21.7%;  Score 444.2;  DB 15;  Length 1531;
  Best Local Similarity   62.8%;  Pred. No. 5.1e-104;
  Matches  689;  Conservative    0;  Mismatches  408;  Indels    0;  Gaps    0;

Qy    250 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 309
          |||| ||||    |||||||||| ||| | || |||    |    || || ||   |||||
Db     29 accctgaagtgaccatgaatataagtcagatgatcacctactggggatacccagctgagg 88

Qy    310 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 369
          |||||||||| |  || ||||| || |||||||||    |  |||| ||||||   ||
Db     89 aatatgaagttgtgaccgaagacggttatatccttgggatcgacagaattccttatggga 148

Qy    370 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 429
             |  |  |||| | |||    || ||||| |  ||  |||| || ||  |  | |
Db    149 ggaaaaattcagagaatataggccggagacctgttgcatttttgcaacacggtttgctcg 208

Qy    430 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 489
           |   || |  |||||||| |||||||||||||||| |||||| |||||| ||||| |
Db    209 catcagccacaaactggatctccaacctgcccaacaacagcctggccttcatcctggccg 268

Qy    490 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 549
          | || || |  ||||||||| |||||||||||||||| ||| ||||| |  | |   |
Db    269 acgccgggtacgacgtgtggctggggaacagcaggggcaacacctgggccaggaggaatc 328

Qy    550 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 609
            |    |||    |||   |  || |||||||||||||| | ||| ||||||||||  |
Db    329 tgtactactcgcccgactccgtcgaattctgggctttcagctttgacgagatggctaaat 388

Qy    610 ttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 669
          |||||||| ||    ||  |||| || ||| |||||||||| ||||| ||| |  | |
Db    389 atgaccttcccgccaccattgacttcatcttgaagaaaacgggacaggacaagctacact 448

Qy    670 atgtcggctattcacagggcaccaccatgggctttattgcatttttccaccatgccagagc 729
          | || ||| |||| |||||||||||||| || || || || ||||||||||  ||  |||
Db    449 acgttggccattcccagggcaccaccattggtttcatcgccttttccaccaatcccaagc 508

Qy    730 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 789
          ||||  |    |||||||  |  | ||| |||||| ||| | ||||| || ||| |  |
Db    509 tggcgaaacggatcaaaaccttctatgcattagctcccgttgccaccgtgaagtacaccg 568

Qy    790 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 849
          ||| ||     | |||| |  || |  | ||     |  |  |||||    | ||||| |
Db    569 aaaccctgttaaacaaactcatgctcgtcccttcgttcctcttcaagcttatatttggaa 628

Qy    850 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccctttgtggcc 909
          | |||  |||       ||   |  |||     |  || | |  . |  | ||   ||
Db    629 acaaaatattctacccacaccacttctttgatcaatttctcgccaccgaggtatgctccc 688

Qy    910 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 969
            | ||   | ||||   | || || ||   | ||||  |  |  |||||||  |||| |
Db    689 gcgagacggtggatctcctctgcagcaacgccctgtttatcatttgtggatttgacacta 748

Qy    970 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 1029
            ||  |||||||||| ||      ||| |||    | || | ||  || |||||||| |
Db    749 tgaacttgaacatgagtcgcttggatgtgtatctgtcacataatccagcaggaacatcgg 808

Qy   1030 tgcaaaatattctacactggagccaggcagtgaatttctggtgaactccgggcatttgact 1089
          | || ||  | || |||||||  |||||| || || |||||  |  |||  || ||||||||
Db    809 ttcagaacgtgctccactggtcccaggctgttaagtctgggaagttccaagcttttgact 868

Qy   1090 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1149
          ||||  ||         | ||  ||    | |   |||||  |  |  |||      ||||
Db    869 ggggaagcccagttcagaacatgatgcactatcatcagagcatgcctccctactacaacc 928

Qy   1150 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1209
          | | ||| |||   || || |  ||| ||||||  || ||  | |||| |||  |  | |
Db    929 tgacagacatgcatgtgccaatcgcagtgtggaacggtggcaacgacttgctggccgacc 988

Qy   1210 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1269
          |  | || ||  |  |  |||| ||  || |  |||| ||||| ||||| | ||| ||||
Db    989 ctcacgatgttgacctttgctttccaagctccccaatctcatttaccacaggaagattc 1048

Qy   1270 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1329
          ||   |    |||| |||| || |||||||  |||||||  |||||     | |||||||
Db    1049 ctcattacaatcacttggactttatctgggccatggatgcccctcaagcggtttacaatg 1108

Qy   1330 aaatcatccatctgatg 1346
          ||||  |    |||||
Db    1109 aaattgtttccatgatg 1125


RESULT  11
AAF28679
ID   AAF28679 standard; cDNA; 699 BP.
XX
AC   AAF28679;
XX
DT   05-APR-2001  (first entry)
XX
DE   Human protein HP03372 coding sequence #1.
XX
KW   Human; hydrophobic domain; immune deficiency; autoimmune disorder;
KW   allergy; tissue growth; regeneration; wound healing; burn; tumour;
KW   periodontal disease; thrombolytic condition; haemostatic condition;
KW   infection; ss.
XX
OS   Homo sapiens.
XX
PN   WO200102563-A2.
XX
PD   11-JAN-2001.
XX
PF   16-JUN-2000; 2000WO-JP03943.
XX
PR   02-JUL-1999;   99JP-0188835.
XX
PA   (SAGA ) SAGAMI CHEM RES CENT.
PA   (PROT-) PROTEGENE INC.
XX
PI   Kato S,  Kimura T;
XX
DR   WPI; 2001-071581/08.
DR   P-PSDB; AAB61608.
XX
PT   New human proteins with hydrophobic domains, useful for the treatment
PT   of immune disorders, tumors, allergic conditions, thrombosis and
PT   microbial infection -
XX
PS   Claim 3; Page 113; 153pp; English.
XX
CC   The present invention relates to human proteins (AAB61608-AAB61617) and
CC   their coding sequences (AAF28679-AAF28698). The proteins of the present
CC   invention have hydrophobic domains and can be used for the treatment of
CC   various immune deficiencies and disorders, such as severe combined
CC   immunodeficiency (SCID), multiple sclerosis, rheumatoid arthritis,
CC   autoimmune pulmonary inflammation, graft-versus-host disease and
CC   Guillain-Barre syndrome. The proteins may also be useful in the treatment
CC   of allergic reactions and conditions, such as asthma and in regulation of
CC   haematopoiesis or lymphoid cell deficiencies. The proteins may also have
```

```
CC   utility in compositions used for bone, cartilage, tendon and/or nerve
CC   tissue growth or regeneration as well as wound healing and in the
CC   treatment of burns. The proteins may be used in the treatment of
CC   periodontal disease and in other tooth repair processes. Other uses
CC   include treatment of thrombolytic and haemostatic conditions, treatment
CC   or prevention of tumours and inhibiting infection by bacteria, viruses,
CC   fungi and other parasites.
XX
SQ   Sequence 699 BP; 216 A; 144 C; 148 G; 191 T; 0 other;


  Query Match             14.0%;  Score 286;  DB 22;  Length 699;
  Best Local Similarity   67.4%;  Pred. No. 1.2e-63;
  Matches  403;  Conservative    0;  Mismatches  195;  Indels    0;  Gaps    0;

Qy    211 gaaatgtgaattcagtacatatgccaactaaagctgtggacccagaagcattcatgaata 270
          ||  | || ||       | | |  |   ||   ||   |||| |||||    |||||||
Db     44 gatctatgtatggttatgacaagaaaggaaacaatgcaaaccctgaagctaatatgaata 103

Qy    271 ttagtgaaatcatccaacatcaaggctatccctgtgaggaatatgaagtcgcaactgaag 330
          ||||  | ||  ||     |    || ||||| | ||| || ||||| ||  ||||  |||
Db    104 ttagccagattatttcttactggggttatccttatgaagagtatgatgttacaacaaaag 163

Qy    331 atgggtatatcctttctgttaacaggattcctcgaggcctagtgcaacctaagaagacag 390
          |||| |||||||||    || | ||||||| | ||   ||     ||   || |||||
Db    164 atggttatatccttggaatttataggattccacatggaagaggatgcccagggaggacag 223

Qy    391 gttccaggcctgtggtgttactgcagcatggcctagttggaggtgctagcaactggattt 450
          | | | |||||  ||||   |||||||||||| || ||| |  ||| || ||||||||||
Db    224 ctccaaagcctgctgtgtatttgcagcatggcttaattgcatctgccagtaactggattt 283

Qy    451 ccaacctgcccaacaatagcctgggcttcattctggcagatgctggttttgacgtgtgga 510
          ||||||||||||||| ||  |||  ||| ||||||||||  ||||| ||||||||||
Db    284 gcaacctgcccaacaacagtttggctttccttctggcagatagtggttatgacgtgtggt 343

Qy    511 tggggaacagcaggggaaacgcctggtctcgaaaacacaagacactctccatagaccaag 570
          |||||||||| | |||||| | |||||  ||||||||   | | | ||    |   |
Db    344 tggggaacagccgaggaaacacttggtccagaaaacaccttaaattgtcaccgaaatcac 403

Qy    571 atgagttctgggctttcagttatgatgagatggctaggtttgaccttcctgcagtgataa 630
            || | |||||| |||||||  ||||||||||||  | |||||||| ||    || |
Db    404 cagaatactgggccttcagtttggatgagatggctaaatatgaccttccagccacaatca 463

Qy    631 actttattttgcagaaaacgggccaggaaaagatctattatgtcggctattcacagggca 690
          | |||||  |  ||||||| || ||| |     |||| || || ||| | ||||| ||||
Db    464 attttatcatagagaaaactggacagaagcgactctactacgtgggccactcacaaggca 523

Qy    691 ccaccatgggctttattgcattttccaccatgccagagctggctcagaaaatcaaaatgt 750
          ||||||| |  ||||| |||||||| || |  ||||| |||||| | || || || || |
Db    524 ccaccatagcttttatagcattttctacaaacccagaactggctaaaaagattaagatat 583

Qy    751 attttgctttagcacccatagccactgttaagcatgcaaaaagccccgggaccaaatt 808
          ||||||  | || ||  | | ||| |||||  |  |  |||| ||   ||  ||| |
Db    584 tttttgcactggctccagttgtcacagttaaatacacccaaagtcctatgaaaaaact 641


RESULT 12
AAF28689
ID   AAF28689 standard; cDNA; 1308 BP.
XX
AC   AAF28689;
XX
DT   05-APR-2001  (first entry)
XX
DE   Human protein HP03372 coding sequence #2.
XX
KW   Human; hydrophobic domain; immune deficiency; autoimmune disorder;
KW   allergy; tissue growth; regeneration; wound healing; burn; tumour;
KW   periodontal disease; thrombolytic condition; haemostatic condition;
KW   infection; ss.
XX
OS   Homo sapiens.
XX
PN   WO200102563-A2.
XX
PD   11-JAN-2001.
XX
PF   16-JUN-2000; 2000WO-JP03943.
XX
PR   02-JUL-1999;   99JP-0188835.
XX
PA   (SAGA ) SAGAMI CHEM RES CENT.
PA   (PROT-) PROTEGENE INC.
XX
PI   Kato S,  Kimura T;
XX
DR   WPI; 2001-071581/08.
DR   P-PSDB; AAB61608.
XX
PT   New human proteins with hydrophobic domains, useful for the treatment
PT   of immune disorders, tumors, allergic conditions, thrombosis and
PT   microbial infection -
XX
PS   Claim 4; Pages 122-124; 153pp; English.
XX
CC   The present invention relates to human proteins (AAB61608-AAB61617) and
CC   their coding sequences (AAF28679-AAF28698). The proteins of the present
CC   invention have hydrophobic domains and can be used for the treatment of
CC   various immune deficiencies and disorders, such as severe combined
CC   immunodeficiency (SCID), multiple sclerosis, rheumatoid arthritis,
CC   autoimmune pulmonary inflammation, graft-versus-host disease and
CC   Guillain-Barre syndrome. The proteins may also be useful in the treatment
CC   of allergic reactions and conditions, such as asthma and in regulation of
CC   haematopoiesis or lymphoid cell deficiencies. The proteins may also have
CC   utility in compositions used for bone, cartilage, tendon and/or nerve
CC   tissue growth or regeneration as well as wound healing and in the
CC   treatment of burns. The proteins may be used in the treatment of
CC   periodontal disease and in other tooth repair processes. Other uses
CC   include treatment of thrombolytic and haemostatic conditions, treatment
CC   or prevention of tumours and inhibiting infection by bacteria, viruses,
CC   fungi and other parasites.
XX
SQ   Sequence 1308 BP; 402 A; 246 C; 264 G; 396 T; 0 other;


  Query Match             14.0%;  Score 286;  DB 22;  Length 1308;
  Best Local Similarity   67.4%;  Pred. No. 1.5e-63;
  Matches  403;  Conservative    0;  Mismatches  195;  Indels    0;  Gaps    0;

Qy    211 gaaatgtgaattcagtacatatgccaactaaagctgtggacccagaagcattcatgaata 270
          ||  | || ||       | | |  |   ||   ||   |||| |||||    |||||||
Db    119 gatctatgtatggttatgacaagaaaggaaacaatgcaaaccctgaagctaatatgaata 178

Qy    271 ttagtgaaatcatccaacatcaaggctatccctgtgaggaatatgaagtcgcaactgaag 330
          ||||  | ||  ||     |    || ||||| | ||| || ||||| ||  ||||  |||
Db    179 ttagccagattatttcttactggggttatccttatgaagagtatgatgttacaacaaaag 238

Qy    331 atgggtatatcctttctgttaacaggattcctcgaggcctagtgcaacctaagaagacag 390
          |||| |||||||||    || | ||||||| | ||   ||     ||   || |||||
Db    239 atggttatatccttggaatttataggattccacatggaagaggatgcccagggaggacag 298

Qy    391 gttccaggcctgtggtgttactgcagcatggcctagttggaggtgctagcaactggattt 450
          | | | |||||  ||||   |||||||||||| || ||| |  ||| || ||||||||||
Db    299 ctccaaagcctgctgtgtatttgcagcatggcttaattgcatctgccagtaactggattt 358

Qy    451 ccaacctgcccaacaatagcctgggcttcattctggcagatgctggttttgacgtgtgga 510
          ||||||||||||||| ||  |||  ||| ||||||||||  ||||| ||||||||||
Db    359 gcaacctgcccaacaacagtttggctttccttctggcagatagtggttatgacgtgtggt 418

Qy    511 tggggaacagcaggggaaacgcctggtctcgaaaacacaagacactctccatagaccaag 570
          |||||||||| | |||||| | |||||  ||||||||   | | | ||    |   |
Db    419 tggggaacagccgaggaaacacttggtccagaaaacaccttaaattgtcaccgaaatcac 478

Qy    571 atgagttctgggctttcagttatgatgagatggctaggtttgaccttcctgcagtgataa 630
            || | |||||| |||||||  ||||||||||||  | |||||||| ||    || |
Db    479 cagaatactgggccttcagtttggatgagatggctaaatatgaccttccagccacaatca 538
```

```
Qy     631 actttatttttgcagaaaacgggccaggaaaagatctattatgtcggctattcacagggca 690
           | |||||  |  ||||||| || ||| |     |||| || || ||| | ||||| ||||
Db     539 attttatcatagagaaaactggacagaagcgactctactacgtgggccactcacaaggca 598

Qy     691 ccaccatgggctttattgcattttccaccatgccagagctggctcagaaaatcaaaatgt 750
           ||||||| |  ||||| |||||||| || |  ||||| |||||| | || || || || |
Db     599 ccaccatagcttttatagcattttctacaaacccagaactggctaaaaagattaagatat 658

Qy     751 attttgctttagcacccatagccactgttaagcatgcaaaaagccccgggaccaaatt 808
           ||||||  | || ||  | | ||| |||||  |  |  |||| ||   ||  ||| |
Db     659 tttttgcactggctccagttgtcacagttaaatacacccaaagtcctatgaaaaaact 716


RESULT 13
AAA44349
ID   AAA44349 standard; cDNA; 617 BP.
XX
AC   AAA44349;
XX
DT   21-AUG-2000  (first entry)
XX
DE   Human secreted expressed sequence tag SEQ ID NO:924.
XX
KW   Human; mouse; chicken; rat; secreted expressed sequence tag; sEST;
KW   expressed sequence tag; EST; probe; chemotactic; proliferative;
KW   immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
KW   thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
KW   antiviral; antidiabetic; antiasthmatic; vulnerary; antiparkinsonian;
KW   antiulcer; osteopathic; neuroprotective; nootropic; antipsoriatic;
KW   cerebroprotective; anticonvulsant; antidepressant; gene therapy;
KW   vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
KW   insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
KW   lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
KW   central nervous system disorder; Alzheimer's disease; stroke;
KW   Parkinson's disease; Huntington's disease; coagulation disorder;
KW   haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
KW   tumour; infection; depression; psoriasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200021991-A1.
XX
PD   20-APR-2000.
XX
PF   15-OCT-1999;   99WO-US24206.
XX
PR   15-OCT-1998;   98US-0104436.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Jacobs K,  McCoy JM,  LaVallie ER,  Collins-Racie LA,  Evans C;
PI   Merberg D,  Treacy M,  Bowman MR;
XX
DR   WPI; 2000-317938/27.
XX
PT   Isolated polynucleotides, and encoded proteins, comprising secreted
PT   expressed sequence tags (sESTs), useful for treating various disorders
PT   such as autoimmune, infectious, and central nervous system disorders -
XX
PS   Claim 1; Page 441; 803pp; English.
XX
CC   AAA43426 to AAA45925 represent specifically claimed secreted expressed
CC   sequence tags (sESTs), isolated from human, mouse, chicken and rat
CC   tissue sources. The sESTs can have a range of activities depending on
CC   the tissues they were isolated from. The activities include:
CC   chemotactic; proliferative; immunomodulatory; haematopoietic;
CC   chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
CC   cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
CC   antiasthmatic; vulnerary; antiulcer; osteopathic; neuroprotective;
CC   nootropic; antiparkinsonian; antipsoriatic; cerebroprotective;
CC   anticonvulsant; and antidepressant. The sESTs can be used for gene
CC   therapy and in vaccines. The sESTs are useful as probes for the
CC   identification and isolation of full-length cDNAs and genomic DNA
CC   molecules which correspond to the sESTs. Proteins encoded by the sESTs
CC   are useful in assays for determining biological activity and raising
CC   antibodies. They may be useful for treatment of autoimmune disorders
CC   (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC   (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
CC   osteoporosis, osteoarthritis, central nervous system disorders
CC   (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC   disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC   disease), tumours, bacterial, fungal or viral infections, depression and
CC   psoriasis. AAA45926 to AAA45931 represent linker variants which are given
CC   in the exemplification of the present invention.
XX
SQ   Sequence 617 BP; 181 A; 125 C; 139 G; 172 T; 0 other;


  Query Match          12.6%;  Score 257.8;  DB 21;  Length 617;
  Best Local Similarity   70.7%;  Pred. No. 1.8e-56;
  Matches  343;  Conservative    0;  Mismatches  142;  Indels    0;  Gaps    0;

Qy     244 ctgtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccct 303
           |||   ||| ||||     ||||| |||||| | || ||     ||   || || ||
Db     133 ctggaagccctgaagtgactatgaacattagtcagatgattacttattggggatacccaa 192

Qy     304 gtgaggaatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctc 363
           ||| ||||||||||| |  |||||||||| ||||| |||   || || || ||||||
Db     193 atgaagaatatgaagttgtgactgaagatggttatattcttgaagtcaatagaattcctt 252

Qy     364 gaggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcc 423
             ||        | |   ||| |||||    || ||||| ||||  ||||||||||
Db     253 atgggaagaaaaattcagggaatacaggccagagacctgttgtgtttttgcagcatggtt 312

Qy     424 tagttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattc 483
           |  || |  || |  ||||||||||||||||||||| ||||| ||||| | ||||||||
Db     313 tgcttgcatcagccacaaactggatttccaacctgccgaacaacagccttgccttcattc 372

Qy     484 tggcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaa 543
           ||||||||||||||| ||| |||||| |||| |||||||| |||||| ||||| |  |||
Db     373 tggcagatgctggttatgatgtgtggctgggcaacagcagaggaaacacctgggccagaa 432

Qy     544 aacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatgg 603
           | ||  |      ||  |||   || ||| ||||||||||||| | |||||| ||||
Db     433 gaaacttgtactattcaccagattcagttgaattctgggctttcagctttgatgaaatgg 492

Qy     604 ctaggtttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaaaga 663
           |||  | |||||||| ||    ||  |||| ||| |  |||||| || ||| |  ||
Db     493 ctaaatatgaccttccagccacaatcgacttcattgtaaagaaaactggacagaagcagc 552

Qy     664 tctattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgc 723
           |  | ||||| ||| |||| |||||||||||||| || |||||||| ||||||||||  |
Db     553 tacactatgttggccattcccagggcaccaccattggttttattgccttttccaccaatc 612

Qy     724 cagag 728
           |||
Db     613 tcgag 617


RESULT 14
AAA43744
ID   AAA43744 standard; cDNA; 572 BP.
XX
AC   AAA43744;
XX
DT   21-AUG-2000  (first entry)
XX
DE   Mouse secreted expressed sequence tag SEQ ID NO:319.
XX
KW   Human; mouse; chicken; rat; secreted expressed sequence tag; sEST;
KW   expressed sequence tag; EST; probe; chemotactic; proliferative;
KW   immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
KW   thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
KW   antiviral; antidiabetic; antiasthmatic; vulnerary; antiparkinsonian;
KW   antiulcer; osteopathic; neuroprotective; nootropic; antipsoriatic;
```

```
KW   cerebroprotective; anticonvulsant; antidepressant; gene therapy;
KW   vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
KW   insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
KW   lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
KW   central nervous system disorder; Alzheimer's disease; stroke;
KW   Parkinson's disease; Huntington's disease; coagulation disorder;
KW   haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
KW   tumour; infection; depression; psoriasis; ss.
XX
OS   Mus musculus.
XX
PN   WO200021991-A1.
XX
PD   20-APR-2000.
XX
PF   15-OCT-1999;   99WO-US24206.
XX
PR   15-OCT-1998;   98US-0104436.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Jacobs K,  McCoy JM,  LaVallie ER,  Collins-Racie LA,  Evans C;
PI   Merberg D,  Treacy M,  Bowman MR;
XX
DR   WPI; 2000-317938/27.
XX
PT   Isolated polynucleotides, and encoded proteins, comprising secreted
PT   expressed sequence tags (sESTs), useful for treating various disorders
PT   such as autoimmune, infectious, and central nervous system disorders -
XX
PS   Claim 1; Page 285-286; 803pp; English.
XX
CC   AAA43426 to AAA45925 represent specifically claimed secreted expressed
CC   sequence tags (sESTs), isolated from human, mouse, chicken and rat
CC   tissue sources. The sESTs can have a range of activities depending on
CC   the tissues they were isolated from. The activities include:
CC   chemotactic; proliferative; immunomodulatory; haematopoietic;
CC   chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
CC   cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
CC   antiasthmatic; vulnerary; antiulcer; osteopathic; neuroprotective;
CC   nootropic; antiparkinsonian; antipsoriatic; cerebroprotective;
CC   anticonvulsant; and antidepressant. The sESTs can be used for gene
CC   therapy and in vaccines. The sESTs are useful as probes for the
CC   identification and isolation of full-length cDNAs and genomic DNA
CC   molecules which correspond to the sESTs. Proteins encoded by the sESTs
CC   are useful in assays for determining biological activity and raising
CC   antibodies. They may be useful for treatment of autoimmune disorders
CC   (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC   (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
CC   osteoporosis, osteoarthritis, central nervous system disorders
CC   (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC   disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC   disease), tumours, bacterial, fungal or viral infections, depression and
CC   psoriasis. AAA45926 to AAA45931 represent linker variants which are given
CC   in the exemplification of the present invention.
XX
SQ   Sequence 572 BP; 179 A; 109 C; 124 G; 156 T; 4 other;

  Query Match             9.5%;  Score 193.2;  DB 21;  Length 572;
  Best Local Similarity   64.1%;  Pred. No. 6.1e-40;
  Matches  288;  Conservative    0;  Mismatches  161;  Indels    0;  Gaps    0;

Qy    248 ggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtga 307
          | | || |||||   ||||||| |||| || || ||  ||||    |  |||||  | ||
Db    123 gaatcctgaagctcacatgaatgttagcgagattattaaacactgggattatccaagcga 182

Qy    308 ggaatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctcgagg 367
          |||||||  || |  |||||  ||||| || || ||| |  |||||  | |||||||  ||
Db    183 anaatatgaggttgtgactgatgatggttacattcttccaattaaccgaattcctcatgg 242

Qy    368 cctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagt 427
                | |||| |  |||  | | |  || || ||  |||||||||  |  |
```

```
Db    243 gaagaacaatgctaatagttcagccccaaagatggtagtattttgtcagcatggcttgct 302

Qy    428 tggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggc 487
          || |    || |     || |||||| | |||   ||| ||||||| ||||||  | ||
Db    303 tgcaacacctggagcaagggtttccaatccgcctgtcaacagcctggccttcatcntagc 362

Qy    488 agatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaaca 547
          |||||||| | ||| |||||||||| | |||||| |||| ||||| |   |||||
Db    363 agatgctgggtatgatgtgtggatgggaagcagcagaggaagtacctgggcaaagaaaca 422

Qy    548 caagacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctag 607
          |   | |||  |   ||    | || ||||||| ||| |||| |||| | |||   |
Db    423 cgtggccctcaacccngattctaaagaattctgggattttagttttgatcaaatgataaa 482

Qy    608 gtttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaaagatcta 667
          | |||||||||| ||    || || || ||| || | ||||| || ||  |  |||| ||
Db    483 atatgaccttccagctaccattaatttcattctggataaaacaggacaaaagcagattta 542

Qy    668 ttatgtcggctattcacagggcaccacca 696
          ||  | ||| |||| || || | | | |
Db    543 ctacattggccattctcaaggaaactcga 571


RESULT  15
AAF58252/c
ID   AAF58252 standard; DNA; 936 BP.
XX
AC   AAF58252;
XX
DT   24-APR-2001  (first entry)
XX
DE   Oligonucleotide D1835.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999;   99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT   a single surface -
XX
PS   Example 6; Page 127; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
XX
SQ   Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

  Query Match             6.2%;  Score 127.6;  DB 22;  Length 936;
  Best Local Similarity   2.7%;  Pred. No. 4.7e-23;
  Matches   16;  Conservative  386;  Mismatches  200;  Indels    0;  Gaps    0;
```

```
Qy   1433 gagggatggggctaggacccatgaaggcagaattacggagagcagagacctagtatacat 1492
          | |||  ||||   ||   |:: ::    : :::::  : :  : : :  :: :::: ::
Db    807 GGGGGGGGGGGGGGGGGGGCWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 748

Qy   1493 ttttcagattccctgcacttggcactaaatccgacacttacatttacatttttttctgt 1552
          :::: : :::   :  : ::   : :::::   : : ::: ::::: :::::::::   :
Db    747 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWGWW 688

Qy   1553 aaattaaagtacttattaggtaaatagaggttttgtatgctattatatattctaccatct 1612
          :::::::: :: ::::::  :::::: :  :::: :::  :::::::::: ::  :: :
Db    687 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 628

Qy   1613 tgaagggtaggttttacctgatagccagaaaatatctagacattctctatatcattcagg 1672
          : ::   ::  :::::  : :::   : ::::::::|:: : ::: : ::::: ::: :
Db    627 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWCWWWWWWWWWWWWWWWWWWWW 568

Qy   1673 taaatctctttaaaacacctattgtttttttctataagccatatttttggagcactaaagt 1732
          ::::: : :::::::: :  :::: :::::: :::::   :::::::  :  : :::: :
Db    567 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 508

Qy   1733 aaaatggcaaattgggacagatattgaggtctggagtctgtggattattgttgactttga 1792
          :::::   :::::|  : : ::::: :  : :  : : : :  :::::: :: : ::: :
Db    507 WWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 448

Qy   1793 caaaataagctagacattttcaccttgttgccacagagacataacactacctcaggaagc 1852
           :::::::  :: : ::::: :  :: ::   : : : : :::: : ::  | || ::
Db    447 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWTAAGCWWWW 388

Qy   1853 tgagctgctttaaggacaacaacaacaaaatcagtgttacagtatggatgaaatctatgt 1912
          : :  :  :::::  : :: :: :: ::::: : : ::: : :::  :: :::: ::: :
Db    387 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 328

Qy   1913 taagcattctcagaataaggccaagttttatagttgcatctcagggaagaaaattttata 1972
          :::  ::: : : :::::    :: :::::::  :  :: : :   :: ::::::::::
Db    327 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWWWWWW 268

Qy   1973 ggatgtttatgagttctccaataaatgcattctgcattacataaaaaaaaaaaaaaaaaa 2032
            :: ::::: : :: :  :::::::  ::: :  :::: ::::::::::::::::::::
Db    267 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 208

Qy   2033 aa 2034
          ::
Db    207 WW 206
```

Search completed: August 24, 2001, 22:41:42
Job time: 7403 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: August 24, 2001, 20:11:58 ; Search time 4395.77 Seconds
(without alignments)
7192.387 Million cell updates/sec

Title: US-09-333-159-45
Perfect score: 2044
Sequence: 1 gtcgacccacgcgtccgggg.....aaaaaaaaaagggcggccgc 2044

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	892	43.6	181170	79	AL353113	AL353113 Homo sapi
c	2	890.4	43.6	158195	63	AC015461	AC015461 Homo sapi
c	3	630.2	30.8	76653	63	AC015506	AC015506 Homo sapi
	4	527.8	25.8	2481	93	HSLYACLY	Z31690 H.sapiens (
	5	527.8	25.8	2481	97	HSU08464	U08464 Human lysos
	6	527.8	25.8	2626	93	HSLAL	X76488 H.sapiens m
	7	526.2	25.7	2493	97	HUMLIPCHL	M74775 Human lysos
	8	489.2	23.9	1378	9	A26689	A26689 Precursor o

c	9	489.2	23.9	1378	9	A26690	A26690 Precursor o
	10	470	23.0	1365	93	HSGLR	X05997 Human mRNA
	11	470	23.0	1367	9	A01046	A01046 H.sapiens m
	12	470	23.0	1367	9	A12714	A12714 Pregastric
	13	467.6	22.9	1336	9	A01157	A01157 R.norvegicu
	14	467.6	22.9	1355	95	RNLIP	X02309 Rat mRNA fo
	15	461	22.6	2358	94	MMLYACLY	Z31689 M.musculus
	16	459.8	22.5	3144	95	S81497	S81497 lysosomal a
	17	458	22.4	1417	7	BOVMPE	L26319 Bos taurus
	18	444.2	21.7	1137	9	AR039022	AR039022 Sequence
	19	444.2	21.7	1137	9	AR092633	AR092633 Sequence
	20	444.2	21.7	1140	9	AR039023	AR039023 Sequence
	21	444.2	21.7	1140	9	AR092634	AR092634 Sequence
	22	444.2	21.7	1146	9	AR039024	AR039024 Sequence
	23	444.2	21.7	1146	9	AR092635	AR092635 Sequence
	24	444.2	21.7	1198	9	A57760	A57760 Sequence 5
	25	444.2	21.7	1528	9	A57756	A57756 Sequence 1
	26	444.2	21.7	1528	9	AR039021	AR039021 Sequence
	27	444.2	21.7	1528	9	AR092632	AR092632 Sequence
	28	444.2	21.7	1651	7	CFLIPASE	Y13899 Canis famil
	29	441	21.6	1137	9	A39301	A39301 Sequence 2
	30	441	21.6	1140	9	A39303	A39303 Sequence 4
	31	441	21.6	1146	9	A39305	A39305 Sequence 6
	32	441	21.6	1528	9	A39300	A39300 Sequence 1
	33	388.8	19.0	1048	9	A57758	A57758 Sequence 3
	34	286	14.0	699	10	AX068249	AX068249 Sequence
	35	286	14.0	1308	10	AX068259	AX068259 Sequence
	36	255	12.5	148720	65	AC019353	AC019353 Homo sapi
	37	255	12.5	161826	90	AL358532	AL358532 Human DNA
	38	255	12.5	172834	62	AC011277	AC011277 Homo sapi
	39	151.2	7.4	76653	63	AC015506	AC015506 Homo sapi
	40	140.6	6.9	401	93	HSLIPCHL04	U04288 Human lysos
	41	140.6	6.9	1851	93	HSLIPA4	X75491 H.sapiens L
c	42	140.6	6.9	180366	81	AL513533	AL513533 Homo sapi
	43	127.6	6.2	40114	6	CELZK6	AF067942 Caenorhab
	44	122.4	6.0	137911	61	AC009769	AC009769 Homo sapi
	45	122.4	6.0	204102	63	AC013826	AC013826 Homo sapi

ALIGNMENTS

```
RESULT   1
AL353113
LOCUS       AL353113   181170 bp    DNA             HTG       08-APR-2001
DEFINITION  Homo sapiens chromosome 10 clone RP11-304I5, *** SEQUENCING IN
            PROGRESS ***, 2 unordered pieces.
ACCESSION   AL353113
VERSION     AL353113.4  GI:13567970
KEYWORDS    HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 181170)
  AUTHORS   Grafham,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 9, 2001 this sequence version replaced gi:9796988.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA304I5
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 180405 bases at least Q40
            Consensus quality: 180692 bases at least Q30
            Consensus quality: 180914 bases at least Q20
            Insert size: 181070; sum-of-contigs
            Insert size: 150393; agarose-fp
            Quality coverage: 7.50x in Q20 bases; sum-of-contigs Quality
            coverage: 9.09x in Q20 bases; agarose-fp
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   174783: contig of 174783 bp in length
            *  174784 174883: gap of      100 bp
            *  174884   181170: contig of 6287 bp in length.
FEATURES             Location/Qualifiers
     source          1. .181170
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-304I5"
                     /clone_lib="RPCI-11.2"
     misc_feature    1. .174783
                     /note="assembly_fragment:02734
                     clone_end:SP6
                     vector_side:left"
     misc_feature    174884. .181170
                     /note="assembly_fragment:02008"
BASE COUNT    57065 a  34281 c  33093 g  56630 t    101 others
ORIGIN


  Query Match           43.6%;  Score 892;  DB 79;  Length 181170;
  Best Local Similarity  100.0%;  Pred. No. 2.2e-213;
  Matches  892;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1126 agccaactcctgtaaggtacagagtcagagatatgacggtccctacagcaatgtggacag 1185
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  134716 AGCCAACTCCTGTAAGGTACAGAGTCAGAGATATGACGGTCCCTACAGCAATGTGGACAG 134775

Qy    1186 gaggtcaggactggctttcaaatccagaagacgtgaaaatgctgctctctgaggtgacca 1245
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  134776 GAGGTCAGGACTGGCTTTCAAATCCAGAAGACGTGAAAATGCTGCTCTCTGAGGTGACCA 134835

Qy    1246 acctcatctaccataagaatattcctgaatgggctcacgtggatttcatctggggtttgg 1305
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  134836 ACCTCATCTACCATAAGAATATTCCTGAATGGGCTCACGTGGATTTCATCTGGGGTTTGG 134895

Qy    1306 atgctcctcaccgtatgtacaatgaaatcatccatctgatgcagcaggaggagaccaacc 1365
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  134896 ATGCTCCTCACCGTATGTACAATGAAATCATCCATCTGATGCAGCAGGAGGAGACCAACC 134955

Qy    1366 tttcccagggacggtgtgaggccgtattgtgaagcatctgacactgacgatcttaggaca 1425
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  134956 TTTCCCAGGGACGGTGTGAGGCCGTATTGTGAAGCATCTGACACTGACGATCTTAGGACA 135015

Qy    1426 acctcctgagggatggggctaggacccatgaaggcagaattacggagagcagagacctag 1485
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  135016 ACCTCCTGAGGGATGGGGCTAGGACCCATGAAGGCAGAATTACGGAGAGCAGAGACCTAG 135075

Qy    1486 tatacatttttcagattccctgcacttggcactaaatccgacacttacatttacatttttt 1545
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  135076 TATACATTTTTCAGATTCCCTGCACTTGGCACTAAATCCGACACTTACATTTACATTTTT 135135

Qy    1546 tttctgtaaattaaagtacttattaggtaaatagaggttttgtatgctattatatattct 1605
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  135136 TTTCTGTAAATTAAAGTACTTATTAGGTAAATAGAGGTTTTGTATGCTATTATATATTCT 135195

Qy    1606 accacttgaagggtaggttttacctgatagccagaaaatatctagacattctctatatc 1665
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db 135196 ACCATCTTGAAGGGTAGGTTTTACCTGATAGCCAGAAAATATCTAGACATTCTCTATATC 135255

Qy   1666 attcaggtaaatctctttaaaacacctattgttttttctataagccatatttttggagca 1725
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 135256 ATTCAGGTAAATCTCTTTAAAACACCTATTGTTTTTTCTATAAGCCATATTTTTGGAGCA 135315

Qy   1726 ctaaagtaaaatggcaaattgggacagatattgaggtctggagtctgtggattattgttg 1785
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 135316 CTAAAGTAAAATGGCAAATTGGGACAGATATTGAGGTCTGGAGTCTGTGGATTATTGTTG 135375

Qy   1786 actttgacaaaataagctagacattttcaccttgttgccacagagacataacactacctc 1845
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 135376 ACTTTGACAAAATAAGCTAGACATTTTCACCTTGTTGCCACAGAGACATAACACTACCTC 135435

Qy   1846 aggaagctgagctgctttaaggacaacaacaacaaaatcagtgttacagtatggatgaaa 1905
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 135436 AGGAAGCTGAGCTGCTTTAAGGACAACAACAACAAAATCAGTGTTACAGTATGGATGAAA 135495

Qy   1906 tctatgttaagcattctcagaataaggccaagttttatagttgcatctcagggaagaaaa 1965
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 135496 TCTATGTTAAGCATTCTCAGAATAAGGCCAAGTTTTATAGTTGCATCTCAGGGAAGAAAA 135555

Qy   1966 ttttataggatgtttatgagttctccaataaatgcattctgcattacataaa 2017
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 135556 TTTTATAGGATGTTTATGAGTTCTCCAATAAATGCATTCTGCATTACATAAA 135607


RESULT   2
AC015461/c
LOCUS       AC015461   158195 bp    DNA             HTG       27-JUL-2000
DEFINITION  Homo sapiens clone RP11-115N17, WORKING DRAFT SEQUENCE, 10
            unordered pieces.
ACCESSION   AC015461
VERSION     AC015461.3  GI:9502445
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 158195)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-115N17
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 158195)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 27, 2000 this sequence version replaced gi:9112670.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L1907
                Center clone name: 115_N_17
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 152466 bases at least Q40
                Consensus quality: 154897 bases at least Q30
                Consensus quality: 156006 bases at least Q20
                Insert size: 157000; agarose-fp
                Insert size: 157295; sum-of-contigs
                Quality coverage: 5.8 in Q20 bases; agarose-fp
                Quality coverage: 5.8 in Q20 bases; sum-of-contigs
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1002: contig of 1002 bp in length
            *     1003  1102: gap of      100 bp
            *     1103     2388: contig of 1286 bp in length
            *     2389  2488: gap of      100 bp
            *     2489     3525: contig of 1037 bp in length
            *     3526  3625: gap of      100 bp
            *     3626    16527: contig of 12902 bp in length
            *    16528  16627: gap of      100 bp
            *    16628    28636: contig of 12009 bp in length
            *    28637  28736: gap of      100 bp
            *    28737    46598: contig of 17862 bp in length
            *    46599  46698: gap of      100 bp
            *    46699    71912: contig of 25214 bp in length
            *    71913  72012: gap of      100 bp
            *    72013    96612: contig of 24600 bp in length
            *    96613  96712: gap of      100 bp
            *    96713   128331: contig of 31619 bp in length
            *   128332  128431: gap of      100 bp
            *   128432   158195: contig of 29764 bp in length.
FEATURES             Location/Qualifiers
     source          1. .158195
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-115N17"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1002
                     /note="assembly_fragment"
     misc_feature    1103. .2388
                     /note="assembly_fragment"
     misc_feature    2489. .3525
                     /note="assembly_fragment"
     misc_feature    3626. .16527
                     /note="assembly_fragment"
     misc_feature    16628. .28636
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    28737. .46598
                     /note="assembly_fragment"
     misc_feature    46699. .71912
                     /note="assembly_fragment"
     misc_feature    72013. .96612
                     /note="assembly_fragment"
     misc_feature    96713. .128331
                     /note="assembly_fragment"
     misc_feature    128432. .158195
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
BASE COUNT    48628 a  30917 c  30188 g  47557 t    905 others
ORIGIN
```

```
  Query Match           43.6%;  Score 890.4;  DB 63;  Length 158195;
  Best Local Similarity  99.9%;  Pred. No. 5.4e-213;
  Matches  891;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    1126 agccaactcctgtaaggtacagagtcagagatatgacggtccctacagcaatgtggacag 1185
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   19482 AGCCAACTCCTGTAAGGTACAGAGTCAGAGATATGACGGTCCCTACAGCAATGTGGACAG 19423

Qy    1186 gaggtcaggactggctttcaaatccagaagacgtgaaaatgctgctctctgaggtgacca 1245
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   19422 GAGGTCAGGACTGGCTTTCAAATCCAGAAGACGTGAAAATGCTGCTCTCTGAGGTGACCA 19363

Qy    1246 acctcatctaccataagaatattcctgaatgggctcacgtggatttcatctggggtttgg 1305
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   19362 ACCTCATCTACCATAAGAATATTCCTGAATGGGCTCACGTGGATTTCATCTGGGGTTTGG 19303

Qy    1306 atgctcctcaccgtatgtacaatgaaatcatccatctgatgcagcaggaggagaccaacc 1365
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   19302 ATGCTCCTCACCGTATGTACAATGAAATCATCCATCTGATGCAGCAGGAGGAGACCAACC 19243

Qy    1366 tttcccagggacggtgtgaggccgtattgtgaagcatctgacactgacgatcttaggaca 1425
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   19242 TTTCCCAGGGACGGTGTGAGGCCGTATTGTGAAGCATCTGACACTGACGATCTTAGGACA 19183

Qy    1426 acctcctgagggatggggctaggacccatgaaggcagaattacggagagcagagacctag 1485
           |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db   19182 ACCTCCTGAGGGATGGGGCTAGGACCCATGAAGGCAGAATTACAGAGAGCAGAGACCTAG 19123

Qy    1486 tatacatttttcagattccctgcacttggcactaaatccgacacttacatttacatttttt 1545
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   19122 TATACATTTTTCAGATTCCCTGCACTTGGCACTAAATCCGACACTTACATTTACATTTTT 19063

Qy    1546 tttctgtaaattaaagtacttattaggtaaatagaggttttgtatgctattatatattct 1605
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   19062 TTTCTGTAAATTAAAGTACTTATTAGGTAAATAGAGGTTTTGTATGCTATTATATATTCT 19003

Qy    1606 accatcttgaagggtaggttttacctgatagccagaaaatatctagacattctctatatc 1665
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   19002 ACCATCTTGAAGGGTAGGTTTTACCTGATAGCCAGAAAATATCTAGACATTCTCTATATC 18943

Qy    1666 attcaggtaaatctctttaaaacacctattgttttttctataagccatatttttggagca 1725
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   18942 ATTCAGGTAAATCTCTTTAAAACACCTATTGTTTTTTCTATAAGCCATATTTTTGGAGCA 18883

Qy    1726 ctaaagtaaaatggcaaattgggacagatattgaggtctggagtctgtggattattgttg 1785
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   18882 CTAAAGTAAAATGGCAAATTGGGACAGATATTGAGGTCTGGAGTCTGTGGATTATTGTTG 18823

Qy    1786 actttgacaaaataagctagacattttcaccttgttgccacagagacataacactacctc 1845
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   18822 ACTTTGACAAAATAAGCTAGACATTTTCACCTTGTTGCCACAGAGACATAACACTACCTC 18763

Qy    1846 aggaagctgagctgctttaaggacaacaacaacaaaatcagtgttacagtatggatgaaa 1905
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   18762 AGGAAGCTGAGCTGCTTTAAGGACAACAACAACAAAATCAGTGTTACAGTATGGATGAAA 18703

Qy    1906 tctatgttaagcattctcagaataaggccaagttttatagttgcatctcagggaagaaaa 1965
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   18702 TCTATGTTAAGCATTCTCAGAATAAGGCCAAGTTTTATAGTTGCATCTCAGGGAAGAAAA 18643

Qy    1966 ttttataggatgtttatgagttctccaataaatgcattctgcattacataaa 2017
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   18642 TTTTATAGGATGTTTATGAGTTCTCCAATAAATGCATTCTGCATTACATAAA 18591


RESULT   3
AC015506/c
LOCUS       AC015506    76653 bp    DNA             HTG       13-JUL-2000
DEFINITION  Homo sapiens clone RP11-21I24, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC015506
VERSION     AC015506.3  GI:9124007
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 76653)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-21I24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 76653)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 13, 2000 this sequence version replaced gi:6468098.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L4028
                Center clone name: 21_I_24
            --------------.
            * NOTE: This record contains 87 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1     795: contig of 795 bp in length
            *      796  895: gap of      100 bp
            *      896    1656: contig of 761 bp in length
            *     1657  1756: gap of      100 bp
            *     1757    2558: contig of 802 bp in length
            *     2559  2658: gap of      100 bp
            *     2659    3434: contig of 776 bp in length
            *     3435  3534: gap of      100 bp
            *     3535    4301: contig of 767 bp in length
            *     4302  4401: gap of      100 bp
            *     4402    5199: contig of 798 bp in length
            *     5200  5299: gap of      100 bp
            *     5300    6085: contig of 786 bp in length
            *     6086  6185: gap of      100 bp
            *     6186    6972: contig of 787 bp in length
            *     6973  7072: gap of      100 bp
            *     7073    7844: contig of 772 bp in length
            *     7845  7944: gap of      100 bp
            *     7945    8731: contig of 787 bp in length
            *     8732  8831: gap of      100 bp
            *     8832    9625: contig of 794 bp in length
            *     9626  9725: gap of      100 bp
            *     9726   10519: contig of 794 bp in length
            *    10520 10619: gap of      100 bp
```

```
*    10620    11417: contig of 798 bp in length
*    11418 11517: gap of      100 bp
*    11518    12296: contig of 779 bp in length
*    12297 12396: gap of      100 bp
*    12397    13180: contig of 784 bp in length
*    13181 13280: gap of      100 bp
*    13281    14065: contig of 785 bp in length
*    14066 14165: gap of      100 bp
*    14166    14949: contig of 784 bp in length
*    14950 15049: gap of      100 bp
*    15050    15822: contig of 773 bp in length
*    15823 15922: gap of      100 bp
*    15923    16718: contig of 796 bp in length
*    16719 16818: gap of      100 bp
*    16819    17621: contig of 803 bp in length
*    17622 17721: gap of      100 bp
*    17722    18517: contig of 796 bp in length
*    18518 18617: gap of      100 bp
*    18618    19390: contig of 773 bp in length
*    19391 19490: gap of      100 bp
*    19491    20253: contig of 763 bp in length
*    20254 20353: gap of      100 bp
*    20354    21134: contig of 781 bp in length
*    21135 21234: gap of      100 bp
*    21235    22028: contig of 794 bp in length
*    22029 22128: gap of      100 bp
*    22129    22905: contig of 777 bp in length
*    22906 23005: gap of      100 bp
*    23006    23787: contig of 782 bp in length
*    23788 23887: gap of      100 bp
*    23888    24648: contig of 761 bp in length
*    24649 24748: gap of      100 bp
*    24749    25529: contig of 781 bp in length
*    25530 25629: gap of      100 bp
*    25630    26414: contig of 785 bp in length
*    26415 26514: gap of      100 bp
*    26515    27308: contig of 794 bp in length
*    27309 27408: gap of      100 bp
*    27409    28169: contig of 761 bp in length
*    28170 28269: gap of      100 bp
*    28270    29059: contig of 790 bp in length
*    29060 29159: gap of      100 bp
*    29160    29953: contig of 794 bp in length
*    29954 30053: gap of      100 bp
*    30054    30834: contig of 781 bp in length
*    30835 30934: gap of      100 bp
*    30935    31716: contig of 782 bp in length
*    31717 31816: gap of      100 bp
*    31817    32590: contig of 774 bp in length
*    32591 32690: gap of      100 bp
*    32691    33471: contig of 781 bp in length
*    33472 33571: gap of      100 bp
*    33572    34334: contig of 763 bp in length
*    34335 34434: gap of      100 bp
*    34435    35215: contig of 781 bp in length
*    35216 35315: gap of      100 bp
*    35316    36083: contig of 768 bp in length
*    36084 36183: gap of      100 bp
*    36184    36962: contig of 779 bp in length
*    36963 37062: gap of      100 bp
*    37063    37806: contig of 744 bp in length
*    37807 37906: gap of      100 bp
*    37907    38682: contig of 776 bp in length
*    38683 38782: gap of      100 bp
*    38783    39570: contig of 788 bp in length
*    39571 39670: gap of      100 bp
*    39671    40463: contig of 793 bp in length
*    40464 40563: gap of      100 bp
*    40564    41353: contig of 790 bp in length
*    41354 41453: gap of      100 bp
*    41454    42216: contig of 763 bp in length
*    42217 42316: gap of      100 bp
*    42317    43093: contig of 777 bp in length
*    43094 43193: gap of      100 bp
*    43194    43966: contig of 773 bp in length
*    43967 44066: gap of      100 bp
*    44067    44841: contig of 775 bp in length
*    44842 44941: gap of      100 bp
*    44942    45707: contig of 766 bp in length
*    45708 45807: gap of      100 bp
*    45808    46598: contig of 791 bp in length
*    46599 46698: gap of      100 bp
*    46699    47457: contig of 759 bp in length
*    47458 47557: gap of      100 bp
*    47558    48363: contig of 806 bp in length
*    48364 48463: gap of      100 bp
*    48464    49263: contig of 800 bp in length
*    49264 49363: gap of      100 bp
*    49364    50154: contig of 791 bp in length
*    50155 50254: gap of      100 bp
*    50255    51047: contig of 793 bp in length
*    51048 51147: gap of      100 bp
*    51148    51939: contig of 792 bp in length
*    51940 52039: gap of      100 bp
*    52040    52823: contig of 784 bp in length
*    52824 52923: gap of      100 bp
*    52924    53714: contig of 791 bp in length
*    53715 53814: gap of      100 bp
*    53815    54604: contig of 790 bp in length
*    54605 54704: gap of      100 bp
*    54705    55491: contig of 787 bp in length
*    55492 55591: gap of      100 bp
*    55592    56385: contig of 794 bp in length
*    56386 56485: gap of      100 bp
*    56486    57264: contig of 779 bp in length
*    57265 57364: gap of      100 bp
*    57365    58144: contig of 780 bp in length
*    58145 58244: gap of      100 bp
*    58245    59001: contig of 757 bp in length
*    59002 59101: gap of      100 bp
*    59102    59875: contig of 774 bp in length
*    59876 59975: gap of      100 bp
*    59976    60753: contig of 778 bp in length
*    60754 60853: gap of      100 bp
*    60854    61636: contig of 783 bp in length
*    61637 61736: gap of      100 bp
*    61737    62508: contig of 772 bp in length
*    62509 62608: gap of      100 bp
*    62609    63382: contig of 774 bp in length
*    63383 63482: gap of      100 bp
*    63483    64260: contig of 778 bp in length

Query Match             30.8%;  Score 630.2;  DB 63;  Length 76653;
Best Local Similarity   99.4%;  Pred. No. 1.4e-147;
Matches  632;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy    1126 agccaactcctgtaaggtacagagtcagagatatgacggtccctacagcaatgtggacag 1185
           ||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||
Db   30724 AGCCAACTCCTGTAAGGTACAGAGTCAGAGATATGACGGTCCTNACAGCAATGTGGACAG 30665

Qy    1186 gaggtcaggactggctttcaaatccagaagacgtgaaaatgctgctctctgaggtgacca 1245
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   30664 GAGGTCAGGACTGGCTTTCAAATCCAGAAGACGTGAAAATGCTGCTCTCTGAGGTGACCA 30605

Qy    1246 acctcatctaccataagaatattcctgaatgggctcacgtggatttcatctggggtttgg 1305
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   30604 ACCTCATCTACCATAAGAATATTCCTGAATGGGCTCACGTGGATTTCATCTGGGGTTTGG 30545

Qy    1306 atgctcctcaccgtatgtacaatgaaatcatccatctgatgcagcaggaggagaccaacc 1365
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   30544 ATGCTCCTCACCGTATGTACAATGAAATCATCCATCTGATGCAGCAGGAGGAGACCAACC 30485

Qy    1366 tttcccagggacggtgtgaggccgtattgtgaagcatctgacactgacgatcttaggaca 1425
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   30484 TTTCCCAGGGACGGTGTGAGGCCGTATTGTGAAGCATCTGACACTGACGATCTTAGGACA 30425
```

```
Qy     1426 acctcctgagggatggggctaggacccatgaaggcagaattacggagagcagagacctag 1485
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    30424 ACCTCCTGAGGGATGGGGCTAGGACCCATGAAGGCAGAATTACGGAGAGCAGAGACCTAG 30365

Qy     1486 tatacattttcagattccctgcacttggcactaaatccgacacttacatttacatttt 1545
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    30364 TATACATTTTCAGATTCCCTGCACTTGGCACTAAATCCGACACTTACATTTACATTTT 30305

Qy     1546 tttctgtaaattaaagtacttattaggtaaatagaggttttgtatgctattatatattct 1605
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    30304 TTTCTGTAAATTAAAGTACTTATTAGGTAAATAGAGGTTTTGTATGCTATTATATATTCT 30245

Qy     1606 accatcttgaagggtaggttttacctgatagccagaaaatatctagacattctctatatc 1665
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    30244 ACCATCTTGAAGGGTAGGTTTTACCTGATAGCCAGAAAATATCTAGACATTCTCTATATC 30185

Qy     1666 attcaggtaaatctctttaaaacacctattgtttttctataagccatattttggagca 1725
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    30184 ATTCAGGTAAATCTCTTTAAAACACCTATTGTTTTTCTATAAGCCATATTTTGGAGCA 30125

Qy     1726 ctaaagtaaaatggcaaattgggacagatattgagg 1761
            ||||||||||||||||||||||||||||||| || ||
Db    30124 CTAAAGTAAAATGGCAAATTGGGACAGATAGTGGGG 30089
```

```
RESULT     4
HSLYACLY
LOCUS       HSLYACLY     2481 bp    mRNA            PRI       17-FEB-1997
DEFINITION  H.sapiens (HepG2) LAL mRNA for lysosomal acid lipase.
ACCESSION   Z31690
VERSION     Z31690.1  GI:506430
KEYWORDS    LAL; lipase; lysosomal acid lipase.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2481)
  AUTHORS   Du,H. and Gregory,G.A.
  TITLE     Structural Conservation of Putative Functional Motifs between Mouse
            and Human Lysosomal Acid Lipase
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2481)
  AUTHORS   Du,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-APR-1994) Hong Du, Division of Human Genetics,
            Children's Hospital Medical Center, 3333 Bernet Street, Cincinnati,
            OH, 45229-3039, USA
FEATURES             Location/Qualifiers
     source          1. .2481
                     /organism="Homo sapiens"
                     /strain="HepG2"
                     /db_xref="taxon:9606"
                     /clone="clone hLAL"
                     /tissue_type="liver"
                     /cell_type="hepatoma"
                     /clone_lib="human liver HepG2 cDNA library/lambda ZAP"
     sig_peptide     34. .96
                     /gene="LAL"
     gene            34. .1233
                     /gene="LAL"
     CDS             34. .1233
                     /gene="LAL"
                     /citation=[1]
                     /codon_start=1
                     /product="lysosomal acid lipase"
                     /protein_id="CAA83495.1"
                     /db_xref="GI:506431"
                     /db_xref="SPTREMBL:Q16529"
                     /translation="MKMRFLGLVVCLVLWTLHSEGSRGKLTAVDPETNMNVSEIISYW
                     GFPSEEYLVETEDGYILCLNRIPHGRKNHSDKGPKPVVFLQHGLLADSSNWVTNLANS
                     SLGFILADAGFDVWMGNSRGNTWSRKHKTLSVSQDEFWAFSYDEMAKYDLPASINFIL
                     NKTGQEQVYYVGHSQGTTIGFIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGR
                     LPDHLIKDLFGDKEFLPQSAFLKWLGTHVCTHVILKELCGNLCFLLCGFNERNLNMSR
                     VDVYTTHSPAGTSVQNMLHWSQAVKFQKFQAFDWGSSAKNYFHYNQSYPPTYNVKDML
                     VPTAVWSGGHDWLADVYDVNILLTQITNLVFHESIPEWEHLDFIWGLDAPWRLYNKII
                     NLMRKYQ"
     variation       79
                     /gene="LAL"
                     /citation=[1]
                     /replace="c"
     mat_peptide     97. .1230
                     /gene="LAL"
                     /product="lysosomal acid lipase"
     variation       100
                     /gene="LAL"
                     /citation=[1]
                     /replace="g"
     polyA_signal    2459. .2464
     polyA_site      2480
BASE COUNT      692 a    477 c    513 g    799 t
ORIGIN
```

```
  Query Match             25.8%;  Score 527.8;  DB 93;  Length 2481;
  Best Local Similarity   65.4%;  Pred. No. 5.3e-122;
  Matches  791;  Conservative    0;  Mismatches  412;  Indels    6;  Gaps    1;

Qy     164 cagaatggaaatgtggcttctgattctggtggcgtatatgttccagagaaatgtgaattc 223
           |||||| ||||| || |  ||      ||||||  | || | |     | |    || ||||
Db      30 CAGAATGAAAATGCGGTTCTTGGGGTTGGTGGTCTGTTTGGTTCTCTGGACCCTGCATTC 89

Qy     224 ag------tacatatgccaactaaagctgtggacccagaagcattcatgaatattagtga 277
            |        |  |  | ||||||||| || ||| ||   |||||| | |||||
Db      90 TGAGGGGTCTAGAGGGAAACTGACAGCTGTGGATCCTGAAACAAACATGAATGTGAGTGA 149

Qy     278 aatcatccaacatcaaggctatccctgtgaggaatatgaagtcgcaactgaagatgggta 337
           ||| |||      |    |  ||   ||||||||||  ||| |    || |||||||| ||
Db     150 AATTATCTCTTACTGGGGATTCCCTAGTGAGGAATACCTAGTTGAGACAGAAGATGGATA 209

Qy     338 tatcctttctgttaacaggattcctcgaggcctagtgcaacctaagaagacaggttccag 397
           ||| || |    ||||| | |||||||  ||          | | | |    | | |||| |||
Db     210 TATTCTGTGCCTTAACCGAATTCCTCATGGGAGGAAGAACCATTCTGACAAAGGTCCCAA 269

Qy     398 gcctgtggtgttactgcagcatggcctagttggaggtgctagcaactggatttccaacct 457
            || || || || ||||| |||||| |   | | || | |||| |||||| |   | |||||
Db     270 ACCAGTTGTCTTCCTGCAACATGGCTTGCTGGCAGATTCTAGTAACTGGGTCACAAACCT 329

Qy     458 gcccaacaatagcctgggcttcattctggcagatgctggtttgacgtgtggatggggaa 517
             ||||| ||||||||||||||||| ||||||||||||||||||||||||||||||||| ||
Db     330 TGCCAACAGCAGCCTGGGCTTCATTCTTGCTGATGCTGGTTTTGACGTGTGGATGGGCAA 389

Qy     518 cagcaggggaaacgcctggtctcgaaaacacaagacactctccatagaccaagatgagtt 577
           |||||| |||||  ||||||||| |||||| ||||||||||  |  | ||  ||||| ||
Db     390 CAGCAGAGGAAATACCTGGTCTCGGAAACATAAGACACTCTCAGTTTCTCAGGATGAATT 449

Qy     578 ctgggctttcagttatgatgagatggctaggtttgacctcctgcagtgataaactttat 637
           |||||||||||||||||||||||||  |  |||||| || ||    || |||||  | || ||
Db     450 CTGGGCTTTCAGTTATGATGAGATGGCAAAATATGACCTACCAGCTTCCATTAACTTCAT 509

Qy     638 tttgcagaaaacgggccaggaaaagatctattatgtcggctattcacagggcaccaccat 697
            | || | ||||| |||||  ||| |  | |||||||| || |||| || |||||||| ||
Db     510 TCTGAATAAAACTGGCCAAGAACAAGTGTATTATGTGGGTCATTCTCAAGGCACCACTAT 569

Qy     698 gggctttattgcattttccaccatgccagagctggctcagaaaatcaaaatgtatttgc 757
           || ||||| ||||||||       || || |||||||| | |   || ||||||| ||||||
Db     570 AGGTTTTATAGCATTTTCACAGATCCCTGAGCTGGCTAAAAGGATTAAAATGTTTTTTGC 629

Qy     758 tttagcacccatagccactgttaagcatgcaaaaagccccgggaccaaattttgttgct 817
            |  |   || | ||  | || |    | |||||    | |||||||
Db     630 CCTGGGTCCTGTGGCTTCCGTCGCCTTCTGTACTAGCCCTATGGCCAAATTAGGACGATT 689

Qy     818 gccagatatgatgatcaaggg attgtttggcaaaaaagaatttctgtatcagaccagatt 877
            |||||   | |||  ||| ||  | |||||  | ||||||||||     ||||    ||
Db     690 ACCAGATCATCTCATTAAGGACTTATTTGGAGACAAAGAATTTCTTCCCCAGAGTGCGTT 749

Qy     878 tctcagacaacttgttatttacctttgtggccaggtgattcttgatcagatttgtagtaa 937
```

```
Db    750 TTTGAAGTGGCTGGGTACCCACGTTTGCACTCATGTCATACTGAAGGAGCTCTGTGGAAA 809

Qy    938 tatcatgttacttctgggtggattcaacaccaacaatatgaacatgagccgagcaagtgt 997
Db    810 TCTCTGTTTTCTTCTGTGTGGATTTAATGAGAGAAATTTAAATATGTCTAGAGTGGATGT 869

Qy    998 atatgctgcccacactcttgctggaacatctgtgcaaaatattctacactggagccaggc 1057
Db    870 ATATACAACACATTCTCCTGCTGGAACTTCTGTGCAAAACATGTTACACTGGAGCCAGGC 929

Qy   1058 agtgaattctggtgaactccgggcatttgactgggggagtgagaccaaaaatctggaaaa 1117
Db    930 TGTTAAATTCCAAAAGTTTCAAGCCTTTGACTGGGGAAGCAGTGCCAAGAATTATTTTCA 989

Qy   1118 atgcaatcagccaactcctgtaaggtacagagtcagagatatgacggtccctacagcaat 1177
Db    990 TTACAACCAGAGTTATCCTCCCACATACAATGTGAAGGACATGCTTGTGCCGACTGCAGT 1049

Qy   1178 gtggacaggaggtcaggactggctttcaaatccagaagacgtgaaaatgctgctctctga 1237
Db   1050 CTGGAGCGGGGGTCACGACTGGCTTGCAGATGTCTACGACGTCAATATCTTACTGACTCA 1109

Qy   1238 ggtgaccaacctcatctaccataagaatattcctgaatgggctcacgtggatttcatctg 1297
Db   1110 GATCACCAACTTGGTGTTCCATGAGAGCATTCCGGAATGGGAGCATCTTGACTTCATTTG 1169

Qy   1298 gggtttggatgctcctcaccgtatgtacaatgaaatcatccatctgatgcagcaggagga 1357
Db   1170 GGGCCTGGATGCCCCTTGGAGGCTTTATAATAAAATTATTAATCTAATGAGGAAATATCA 1229

Qy   1358 gaccaacct 1366
Db   1230 GTGAAAGCT 1238
```

```
RESULT   5
HSU08464
LOCUS       HSU08464     2481 bp    mRNA            PRI       23-JUN-1994
DEFINITION  Human lysosomal acid lipase mRNA, complete cds.
ACCESSION   U08464
VERSION     U08464.1  GI:505052
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2481)
  AUTHORS   Du,H. and Gregory,G.A.
  TITLE     Structural Conservation of Putative Functional Motifs between Mouse
            and Human Lysosomal Acid Lipase
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2481)
  AUTHORS   Du,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-APR-1994) Hong Du, Division of Human Gentics,
            Children's Hospital Medical Center, 3333 Bernet Street, Cincinnati,
            OH 45229-3039, USA
FEATURES             Location/Qualifiers
     source          1. .2481
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="clone hLAL"
                     /clone_lib="HepG2 cDNA library/Lambda ZAP"
                     /cell_line="HepG2"
                     /cell_type="hepatoma"
                     /tissue_type="liver"
     5'UTR           1. .33
     sig_peptide     34. .96
     CDS             34. .1233
                     /codon_start=1
                     /product="lysosomal acid lipase"
                     /protein_id="AAB60328.1"
                     /db_xref="GI:505053"
                     /translation="MKMRFLGLVVCLVLWTLHSEGSRGKLTAVDPETNMNVSEIISYW
                     GFPSEEYLVETEDGYILCLNRIPHGRKNHSDKGPKPVVFLQHGLLADSSNWVTNLANS
                     SLGFILADAGFDVWMGNSRGNTWSRKHKTLSVSQDEFWAFSYDEMAKYDLPASINFIL
                     NKTGQEQVYYVGHSQGTTIGFIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGR
                     LPDHLIKDLFGDKEFLPQSAFLKWLGTHVCTHVILKELCGNLCFLLCGFNERNLNMSR
                     VDVYTTHSPAGTSVQNMLHWSQAVKFQKFQAFDWGSSAKNYFHYNQSYPPTYNVKDML
                     VPTAVWSGGHDWLADVYDVNILLTQITNLVFHESIPEWEHLDFIWGLDAPWRLYNKII
                     NLMRKYQ"
     variation       79
                     /replace="c"
     mat_peptide     97. .1230
                     /product="lysosomal acid lipase"
     variation       100
                     /replace="g"
     3'UTR           1234. .2480
     polyA_signal    2459. .2464
     polyA_site      2480
BASE COUNT      692 a    477 c    513 g    799 t
ORIGIN
```

Query Match 25.8%; Score 527.8; DB 97; Length 2481;
Best Local Similarity 65.4%; Pred. No. 5.3e-122;
Matches 791; Conservative 0; Mismatches 412; Indels 6; Gaps 1;

```
Qy    164 cagaatggaaatgtggcttctgattctggtggcgtatatgttccagagaaatgtgaattc 223
Db     30 CAGAATGAAAATGCGGTTCTTGGGGTTGGTGGTCTGTTTGGTTCTCTGGACCCTGCATTC 89

Qy    224 ag------tacatatgccaactaaagctgtggacccagaagcattcatgaatattagtga 277
Db     90 TGAGGGGTCTAGAGGGAAACTGACAGCTGTGGATCCTGAAACAAACATGAATGTGAGTGA 149

Qy    278 aatcatccaacatcaaggctatccctgtgaggaatatgaagtcgcaactgaagatgggta 337
Db    150 AATTATCTCTTACTGGGGATTCCCTAGTGAGGAATACCTAGTTGAGACAGAAGATGGATA 209

Qy    338 tatcctttctgttaacaggattcctcgaggcctagtgcaacctaagaagacaggttccag 397
Db    210 TATTCTGTGCCTTAACCGAATTCCTCATGGGAGGAAGAACCATTCTGACAAAGGTCCCAA 269

Qy    398 gcctgtggtgttactgcagcatggcctagttggaggtgctagcaactggatttccaacct 457
Db    270 ACCAGTTGTCTTCCTGCAACATGGCTTGCTGGCAGATTCTAGTAACTGGGTCACAAACCT 329

Qy    458 gcccaacaatagcctgggcttcattctggcagatgctggttttgacgtgtggatggggaa 517
Db    330 TGCCAACAGCAGCCTGGGCTTCATTCTTGCTGATGCTGGTTTTGACGTGTGGATGGGCAA 389

Qy    518 cagcaggggaaacgcctggtctcgaaaacacaagacactctccatagaccaagatgagtt 577
Db    390 CAGCAGAGGAAATACCTGGTCTCGGAAACATAAGACACTCTCAGTTTCTCAGGATGAATT 449

Qy    578 ctgggctttcagttatgatgagatggctaggtttgaccttcctgcagtgataaactttat 637
Db    450 CTGGGCTTTCAGTTATGATGAGATGGCAAAATATGACCTACCAGCTTCCATTAACTTCAT 509

Qy    638 tttgcagaaaacgggccaggaaaagatctattatgtcggctattcacagggcaccaccat 697
Db    510 TCTGAATAAAACTGGCCAAGAACAAGTGTATTATGTGGGTCATTCTCAAGGCACCACTAT 569

Qy    698 gggctttattgcattttccaccatgccagagctggctcagaaaatcaaaatgtattttgc 757
Db    570 AGGTTTTATAGCATTTTCACAGATCCCTGAGCTGGCTAAAAGGATTAAAATGTTTTTTGC 629

Qy    758 tttagcacccatagccactgttaagcatgcaaaaagccccgggaccaaatttttgttgct 817
Db    630 CCTGGGTCCTGTGGCTTCCGTCGCCTTCTGTACTAGCCCTATGGCCAAATTAGGACGATT 689

Qy    818 gccagatatgatgatcaagggattgtttggcaaaaaagaatttctgtatcagaccagatt 877
Db    690 ACCAGATCATCTCATTAAGGACTTATTTGGAGACAAAGAATTTCTTCCCCAGAGTGCGTT 749
```

```
Qy     878 tctcagacaacttgttatttacctttgtggccaggtgattcttgatcagatttgtagtaa 937
           | | |     || | ||   || ||||    || || || ||   |  || |.||| | ||
Db     750 TTTGAAGTGGCTGGGTACCCACGTTTGCACTCATGTCATACTGAAGGAGCTCTGTGGAAA 809

Qy     938 tatcatgttacttctgggtggattcaacaccaacaatatgaacatgagccgagcaagtgt 997
           | ||   || |||||| |||||||| ||    |  ||| | || |||    |||   |||
Db     810 TCTCTGTTTTCTTCTGTGTGGATTTAATGAGAGAAATTTAAATATGTCTAGAGTGGATGT 869

Qy     998 atatgctgcccacactcttgctggaacatctgtgcaaaatattctacactggagccaggc 1057
           |||| |  | ||  ||| |||||||||| |||||||||| ||  ||||||||||||||||
Db     870 ATATACAACACATTCTCCTGCTGGAACTTCTGTGCAAAACATGTTACACTGGAGCCAGGC 929

Qy    1058 agtgaattctggtgaactccgggcatttgactgggggagtgagaccaaaaatctggaaaa 1117
           || || |      | | |  || |||||||||| ||     |||| |||       |
Db     930 TGTTAAATTCCAAAAGTTTCAAGCCTTTGACTGGGGAAGCAGTGCCAAGAATTATTTTCA 989

Qy    1118 atgcaatcagccaactcctgtaaggtacagagtcagagatatgacggtccctacagcaat 1177
           | ||| |||     |||| | ||||  || |  || |||  || || || ||| |
Db     990 TTACAACCAGAGTTATCCTCCCACATACAATGTGAAGGACATGCTTGTGCCGACTGCAGT 1049

Qy    1178 gtggacaggaggtcaggactggctttcaaatccagaagacgtgaaaatgctgctctctga 1237
           ||||  || ||||| |||||||||| || ||    | ||||| || ||  | ||  || |
Db    1050 CTGGAGCGGGGGTCACGACTGGCTTGCAGATGTCTACGACGTCAATATCTTACTGACTCA 1109

Qy    1238 ggtgaccaacctcatctaccataagaatattcctgaatgggctcacgtggatttcatctg 1297
           | | |||||| |  | | |||| |||  ||||| |||||||  ||  | || |||||  ||
Db    1110 GATCACCAACTTGGTGTTCCATGAGAGCATTCCGGAATGGGAGCATCTTGACTTCATTTG 1169

Qy    1298 gggtttggatgctcctcaccgtatgtacaatgaaatcatccatctgatgcagcaggagga 1357
           |||  ||||||| |||    |  | || ||| |||| ||  |||| ||| | |  | |
Db    1170 GGGCCTGGATGCCCCTTGGAGGCTTTATAATAAAATTATTAATCTAATGAGGAAATATCA 1229

Qy    1358 gaccaacct 1366
           |   || ||
Db    1230 GTGAAAGCT 1238
```

```
RESULT   6
HSLAL
LOCUS       HSLAL        2626 bp    mRNA            PRI       25-FEB-1994
DEFINITION  H.sapiens mRNA for lysosomal acid lipase.
ACCESSION   X76488
VERSION     X76488.1  GI:434305
KEYWORDS    lysosomal acid lipase.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2626)
  AUTHORS   Ameis,D., Merkel,M., Eckerskorn,C. and Greten,H.
  TITLE     Purification, characterization and molecular cloning of human
            hepatic lysosomal acid lipase
  JOURNAL   Eur. J. Biochem. 219 (3), 905-914 (1994)
  MEDLINE   94155897
REFERENCE   2  (bases 1 to 2626)
  AUTHORS   Ameis,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-NOV-1993) D. Ameis, Medical Department, University
            Hospital Eppendorf, Martinistrasse 52, 20246 Hamburg, FRG
FEATURES             Location/Qualifiers
     source          1. .2626
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /tissue_type="liver"
     gene            146. .1345
                     /gene="LAL/LIPA"
     CDS             146. .1345
                     /gene="LAL/LIPA"
                     /EC_number="3.1.1.13"
                     /note="sterol esterase"
                     /codon_start=1
                     /evidence=experimental
                     /product="lysosomal acid lipase"
                     /protein_id="CAA54026.1"
                     /db_xref="GI:434306"
                     /db_xref="SWISS-PROT:P38571"
                     /translation="MKMRFLGLVVCLVLWTLHSEGSGGKLTAVDPETNMNVSEIISYW
                     GFPSEEYLVETEDGYILCLNRIPHGRKNHSDKGPKPVVFLQHGLLADSSNWVTNLANS
                     SLGFILADAGFDVWMGNSRGNTWSRKHKTLSVSQDEFWAFSYDEMAKYDLPASINFIL
                     NKTGQEQVYYVGHSQGTTIGFIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGR
                     LPDHLIKDLFGDKEFLPQSAFLKWLGTHVCTHVILKELCGNLCFLLCGFNERNLNMSR
                     VDVYTTHSPAGTSVQNMLHWSQAVKFQKFQAFDWGSSAKNYFHYNQSYPPTYNVKDML
                     VPTAVWSGGHDWLADVYDVNILLTQITNLVFHESIPEWEHLDFIWGLDAPWRLYNKII
                     NLMRKYQ"
     sig_peptide     146. .226
                     /gene="LAL/LIPA"
                     /evidence=experimental
     misc_feature    227. .373
                     /gene="LAL/LIPA"
                     /note="propeptide"
                     /evidence=experimental
     mat_peptide     374. .1342
                     /gene="LAL/LIPA"
                     /evidence=experimental
                     /product="lysosomal acid lipase"
BASE COUNT      741 a    527 c    547 g    811 t
ORIGIN
```

Query Match 25.8%; Score 527.8; DB 93; Length 2626;
Best Local Similarity 65.4%; Pred. No. 5.3e-122;
Matches 791; Conservative 0; Mismatches 412; Indels 6; Gaps 1;

```
Qy     164 cagaatggaaatgtggcttctgattctggtggcgtatatgttccagagaaatgtgaattc 223
           ||||||| ||||| || |  ||    ||||||  | | || | |   | |   || ||||
Db     142 CAGAATGAAAATGCGGTTCTTGGGGTTGGTGGTCTGTTTGGTTCTCTGGACCCTGCATTC 201

Qy     224 ag------tacatatgccaactaaagctgtggacccagaagcattcatgaatattagtga 277
           |       |  |   | ||||||||| || ||| ||  |||||||  | ||||||
Db     202 TGAGGGGTCTGGAGGGAAACTGACAGCTGTGGATCCTGAAACAAACATGAATGTGAGTGA 261

Qy     278 aatcatccaacatcaaggctatccctgtgaggaatatgaagtcgcaactgaagatgggta 337
           ||| |||    |    || |  ||  ||||||||||   ||| |  || ||||||||  ||
Db     262 AATTATCTCTTACTGGGGATTCCCTAGTGAGGAATACCTAGTTGAGACAGAAGATGGATA 321

Qy     338 tatcctttctgttaacaggattcctcgaggcctagtgcaacctaagaagacaggttccag 397
           ||| || |   ||||| | |||||||  ||       | |.| |    | | |||| |||
Db     322 TATTCTGTGCCTTAACCGAATTCCTCATGGGAGGAAGAACCATTCTGACAAAGGTCCCAA 381

Qy     398 gcctgtggtgttactgcagcatggcctagttggaggtgctagcaactggatttccaacct 457
           || || || || ||||| |||||| |  | | || | |||| |||||| |  | |||||
Db     382 ACCAGTTGTCTTCCTGCAACATGGCTTGCTGGCAGATTCTAGTAACTGGGTCACAAACCT 441

Qy     458 gcccaacaatagcctgggcttcattctggcagatgctggttttgacgtgtggatggggaa 517
           ||||||  |||||||||||||||||| || ||||||||||||||||||||||||||||||
Db     442 TGCCAACAGCAGCCTGGGCTTCATTCTTGCTGATGCTGGTTTTGACGTGTGGATGGGCAA 501

Qy     518 cagcaggggaaacgcctggtctcgaaaacacaagacactctccatagaccaagatgagtt 577
           |||||| |||||  |||||||||| ||||| ||||||||||  |    || |||||  ||
Db     502 CAGCAGAGGAAATACCTGGTCTCGGAAACATAAGACACTCTCAGTTTCTCAGGATGAATT 561

Qy     578 ctgggctttcagttatgatgagatggctaggtttgaccttcctgcagtgataaactttat 637
           |||||||||||||||||||||||||||| |  | |||||| || ||    || ||||| ||
Db     562 CTGGGCTTTCAGTTATGATGAGATGGCAAAATATGACCTACCAGCTTCCATTAACTTCAT 621

Qy     638 tttgcagaaaacgggccaggaaaagatctattatgtcggctattcacagggcaccaccat 697
           | || | ||||| |||| ||| |  | |||||||| ||  |||| || ||||||||  ||
Db     622 TCTGAATAAAACTGGCCAAGAACAAGTGTATTATGTGGGTCATTCTCAAGGCACCACTAT 681

Qy     698 gggctttattgcattttccaccatgccagagctggctcagaaaatcaaaatgtattttgc 757
           || ||||| |||||||||    || || |||||||||| | |  || |||||||  ||||||
Db     682 AGGTTTTATAGCATTTTCACAGATCCCTGAGCTGGCTAAAAGGATTAAAATGTTTTTTGC 741

Qy     758 tttagcacccatagccactgttaagcatgcaaaaagccccgggaccaaatttttgttgct 817
```

```
Db    742 CCTGGGTCCTGTGGCTTCCGTCGCCTTCTGTACTAGCCCTATGGCCAAATTAGGACGATT 801

Qy    818 gccagatatgatgatcaagggattgtttggcaaaaaagaatttctgtatcagaccagatt 877
Db    802 ACCAGATCATCTCATTAAGGACTTATTTGGAGACAAAGAATTTCTTCCCCAGAGTGCGTT 861

Qy    878 tctcagacaacttgttatttacctttgtggccaggtgattcttgatcagatttgtagtaa 937
Db    862 TTTGAAGTGGCTGGGTACCCACGTTTGCACTCATGTCATACTGAAGGAGCTCTGTGGAAA 921

Qy    938 tatcatgttacttctgggtggattcaacaccaacaatatgaacatgagccgagcaagtgt 997
Db    922 TCTCTGTTTTCTTCTGTGTGGATTTAATGAGAGAAATTTAAATATGTCTAGAGTGGATGT 981

Qy    998 atatgctgcccacactcttgctggaacatctgtgcaaaatattctacactggagccaggc 1057
Db    982 ATATACAACACATTCTCCTGCTGGAACTTCTGTGCAAAACATGTTACACTGGAGCCAGGC 1041

Qy   1058 agtgaattctggtgaactccgggcatttgactgggggagtgagaccaaaaatctggaaaa 1117
Db   1042 TGTTAAATTCCAAAAGTTTCAAGCCTTTGACTGGGGAAGCAGTGCCAAGAATTATTTTCA 1101

Qy   1118 atgcaatcagccaactcctgtaaggtacagagtcagagatatgacggtccctacagcaat 1177
Db   1102 TTACAACCAGAGTTATCCTCCCACATACAATGTGAAGGACATGCTTGTGCCGACTGCAGT 1161

Qy   1178 gtggacaggaggtcaggactggctttcaaatccagaagacgtgaaaatgctgctctctga 1237
Db   1162 CTGGAGCGGGGGTCACGACTGGCTTGCAGATGTCTACGACGTCAATATCTTACTGACTCA 1221

Qy   1238 ggtgaccaacctcatctaccataagaatattcctgaatgggctcacgtggatttcatctg 1297
Db   1222 GATCACCAACTTGGTGTTCCATGAGAGCATTCCGGAATGGGAGCATCTTGACTTCATTTG 1281

Qy   1298 gggtttggatgctcctcaccgtatgtacaatgaaatcatccatctgatgcagcaggagga 1357
Db   1282 GGGCCTGGATGCCCCTTGGAGGCTTTATAATAAAATTATTAATCTAATGAGGAAATATCA 1341

Qy   1358 gaccaacct 1366
Db   1342 GTGAAAGCT 1350
```

```
RESULT   7
HUMLIPCHL
LOCUS       HUMLIPCHL    2493 bp    mRNA            PRI       07-JAN-1995
DEFINITION  Human lysosomal acid lipase/cholesteryl esterase mRNA, complete
            cds.
ACCESSION   M74775
VERSION     M74775.1  GI:187151
KEYWORDS    lysosomal acid lipase/cholesteryl esterase.
SOURCE      Homo sapiens fibroblast cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2493)
  AUTHORS   Anderson,R.A. and Sando,G.N.
  TITLE     Cloning and expression of cDNA encoding human lysosomal acid
            lipase/cholesteryl ester hydrolase. Similarities to gastric and
            lingual lipases
  JOURNAL   J. Biol. Chem. 266 (33), 22479-22484 (1991)
  MEDLINE   92042192
FEATURES             Location/Qualifiers
     source          1. .2493
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="fibroblast"
                     /map="10"
     mRNA            <1. .2493
                     /gene="LIPA"
                     /note="G00-120-153"
     gene            1. .2493
                     /gene="LIPA"
     CDS             41. .1240
                     /gene="LIPA"
                     /EC_number="3.1.1.13"
                     /note="alternate initiator ATG may be at position 47"
                     /codon_start=1
                     /db_xref="GDB:G00-120-153"
                     /product="lysosomal acid lipase/cholesteryl esterase"
                     /protein_id="AAA59519.1"
                     /db_xref="GI:187152"
                     /translation="MKMRFLGLVVCLVLWPLHSEGSGGKLTAVDPETNMNVSEIISYW
                     GFPSEEYLVETEDGYILCLNRIPHGRKNHSDKGPKPVVFLQHGLLADSSNWVTNLANS
                     SLGFILADAGFDVWMGNSRGNTWSRKHKTLSVSQDEFWAFSYDEMAKYDLPASINFIL
                     NKTGQEQVYYVGHSQGTTIGFIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGR
                     LPDHLIKDLFGDKEFLPQSAFLKWLGTHVCTHVILKELCGNLCFLLCGFNERNLNMSR
                     VDVYTTHSPAGTSVQNMLHWSQAVKFQKFQAFDWGSSAKNYFHYNQSYPPTYNVKDML
                     VPTAVWSGGHDWLADVYDVNILLTQITNLVFHESIPEWEHLDFIWGLDAPWRLYNKII
                     NLMRKYQ"
     sig_peptide     41. .103
                     /gene="LIPA"
                     /note="alternate signal peptide may begin at position 47;
                     as predicted from von Heijne, G. (1986) Nucl. Acids Res.
                     14,4683-4690; G00-120-153"
     mat_peptide     104. .1237
                     /gene="LIPA"
                     /EC_number="3.1.1.13"
                     /note="G00-120-153"
                     /product="lysosomal acid lipase/cholesteryl esterase"
     polyA_signal    2466. .2471
                     /gene="LIPA"
                     /note="G00-120-153"
     polyA_site      2493
                     /gene="LIPA"
                     /note="G00-120-153"
BASE COUNT      692 a    484 c    516 g    801 t
ORIGIN
```

```
  Query Match          25.7%;  Score 526.2;  DB 97;  Length 2493;
  Best Local Similarity   65.3%;  Pred. No. 1.3e-121;
  Matches  790;  Conservative    0;  Mismatches  413;  Indels    6;  Gaps    1;

Qy    164 cagaatggaaatgtggcttctgattctggtggcgtatatgttccagagaaatgtgaattc 223
Db     37 CAGAATGAAAATGCGGTTCTTGGGGTTGGTGGTCTGTTTGGTTCTCTGGCCCCTGCATTC 96

Qy    224 ag------tacatatgccaactaaagctgtggacccagaagcattcatgaatattagtga 277
Db     97 TGAGGGGTCTGGAGGGAAACTGACAGCTGTGGATCCTGAAACAAACATGAATGTGAGTGA 156

Qy    278 aatcatccaacatcaaggctatccctgtgaggaatatgaagtcgcaactgaagatgggta 337
Db    157 AATTATCTCTTACTGGGGATTCCCTAGTGAGGAATACCTAGTTGAGACAGAAGATGGATA 216

Qy    338 tatcctttctgttaacaggattcctcgaggcctagtgcaacctaagaagacaggttccag 397
Db    217 TATTCTGTGCCTTAACCGAATTCCTCATGGGAGGAAGAACCATTCTGACAAAGGTCCCAA 276

Qy    398 gcctgtggtgttactgcagcatggcctagttggaggtgctagcaactggatttccaacct 457
Db    277 ACCAGTTGTCTTCCTGCAACATGGCTTGCTGGCAGATTCTAGTAACTGGGTCACAAACCT 336

Qy    458 gcccaacaatagcctgggcttcattctggcagatgctggttttgacgtgtggatggggaa 517
Db    337 TGCCAACAGCAGCCTGGGCTTCATTCTTGCTGATGCTGGTTTTGACGTGTGGATGGGCAA 396

Qy    518 cagcaggggaaacgcctggtctcgaaaacacaagacactctccatagaccaagatgagtt 577
Db    397 CAGCAGAGGAAATACCTGGTCTCGGAAACATAAGACACTCTCAGTTTCTCAGGATGAATT 456

Qy    578 ctgggctttcagttatgatgagatggctaggtttgaccttcctgcagtgataaactttat 637
Db    457 CTGGGCTTTCAGTTATGATGAGATGGCAAAATATGACCTACCAGCTTCCATTAACTTCAT 516
```

```
Qy    638 tttgcagaaaacgggccaggaaaagatctattatgtcggctattcacagggcaccaccat 697
Db    517 TCTGAATAAAACTGGCCAAGAACAAGTGTATTATGTGGGTCATTCTCAAGGCACCACTAT 576

Qy    698 gggctttattgcattttccaccatgccagagctggctcagaaaatcaaaatgtatttttgc 757
Db    577 AGGTTTTATAGCATTTTCACAGATCCCTGAGCTGGCTAAAAGGATTAAAATGTTTTTTGC 636

Qy    758 tttagcacccatagccactgttaagcatgcaaaaagccccgggaccaaattttttgttgct 817
Db    637 CCTGGGTCCTGTGGCTTCCGTCGCCTTCTGTACTAGCCCTATGGCCAAATTAGGACGATT 696

Qy    818 gccagatatgatgatcaagggattgtttggcaaaaaagaatttctgtatcagaccagatt 877
Db    697 ACCAGATCATCTCATTAAGGACTTATTTGGAGACAAAGAATTTCTTCCCCAGAGTGCGTT 756

Qy    878 tctcagacaacttgttatttacctttgtggccaggtgattcttgatcagatttgtagtaa 937
Db    757 TTTGAAGTGGCTGGGTACCCACGTTTGCACTCATGTCATACTGAAGGAGCTCTGTGGAAA 816

Qy    938 tatcatgttacttctgggtggattcaacaccaacaatatgaacatgagccgagcaagtgt 997
Db    817 TCTCTGTTTTCTTCTGTGTGGATTTAATGAGAGAAATTTAAATATGTCTAGAGTGGATGT 876

Qy    998 atatgctgcccacactcttgctggaacatctgtgcaaaatattctacactggagccaggc 1057
Db    877 ATATACAACACATTCTCCTGCTGGAACTTCTGTGCAAAACATGTTACACTGGAGCCAGGC 936

Qy   1058 agtgaattctggtgaactccgggcatttgactggggggagtgagaccaaaaatctggaaaa 1117
Db    937 TGTTAAATTCCAAAAGTTTCAAGCCTTTGACTGGGGAAGCAGTGCCAAGAATTATTTTCA 996

Qy   1118 atgcaatcagccaactcctgtaaggtacagagtcagagatatgacggtccctacagcaat 1177
Db    997 TTACAACCAGAGTTATCCTCCCACATACAATGTGAAGGACATGCTTGTGCCGACTGCAGT 1056

Qy   1178 gtggacaggaggtcaggactggctttcaaatccagaagacgtgaaaatgctgctctctga 1237
Db   1057 CTGGAGCGGGGGTCACGACTGGCTTGCAGATGTCTACGACGTCAATATCTTACTGACTCA 1116

Qy   1238 ggtgaccaacctcatctaccataagaatattcctgaatgggctcacgtggatttcatctg 1297
Db   1117 GATCACCAACTTGGTGTTCCATGAGAGCATTCCGGAATGGGAGCATCTTGACTTCATTTG 1176

Qy   1298 gggtttggatgctcctcaccgtatgtacaatgaaatcatccatctgatgcagcaggagga 1357
Db   1177 GGGCCTGGATGCCCCTTGGAGGCTTTATAATAAAATTATTAATCTAATGAGGAAATATCA 1236

Qy   1358 gaccaacct 1366
Db   1237 GTGAAAGCT 1245
```

```
RESULT    8
A26689
LOCUS       A26689       1378 bp    DNA              PAT       05-APR-1995
DEFINITION  Precursor of rabit gastric lipase coding sequence.
ACCESSION   A26689
VERSION     A26689.1  GI:905029
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 1378)
  AUTHORS   Benicourt,C., Blanchard,C. and Junien,J.L.
  TITLE     Recombinant gastric lipase from rabbit and pharmaceutical
            compositions
  JOURNAL   Patent: EP 0542629-A 9 19-MAY-1993;
            INSTITUT DE RECHERCHE JOUVEINAL
FEATURES             Location/Qualifiers
     source          1..1378
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
BASE COUNT      397 a    287 c    276 g    418 t
ORIGIN

  Query Match             23.9%;  Score 489.2;  DB 9;  Length 1378;
  Best Local Similarity   65.0%;  Pred. No. 2.5e-112;
  Matches  722;  Conservative    0;  Mismatches  388;  Indels    0;  Gaps    0;

Qy    250 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 309
Db    136 ACCCTGAAGTGAATATGAATATTAGTCAGATGATTTCCTACTGGGGATACCCAAGTGAAA 195

Qy    310 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 369
Db    196 AATATGAAGTTGTGACTGAAGATGGTTATATCCTTGAGGTTAACAGAATTCCTTATGGGA 255

Qy    370 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 429
Db    256 AGAAAAATTCAGGGAACAGAGGCCAGAGACCCGTCGTATTTTTGCAGCATGGTTTGCTTG 315

Qy    430 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 489
Db    316 CATCAGCCTCAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTTATTCTGGCAG 375

Qy    490 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 549
Db    376 ATGCTGGTTATGGTGTGTGGCTGGGGAACAGCAGAGGAAATACCTGGTCCAGAAGAAACT 435

Qy    550 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 609
Db    436 TATACTATTCACCAGACTCCGTTGAATTCTGGGCTTTTAGCTTTGATGAAATGGCTAAAT 495

Qy    610 ttgaccttcctgcagtgataaactttatttgcagaaaacgggccaggaaaagatctatt 669
Db    496 ATGACCTTCCAGCCACAATTGACTTCATTGTAAAGGAAACTGGACAGGAAAAGCTGCACT 555

Qy    670 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 729
Db    556 ATGTTGGTCATTCCCAGGGCACCACCATTGGTTTTATTGCCTTTTCTACAAATCCCAAGC 615

Qy    730 tggctcagaaaatcaaaatgtatttttgctttagcacccatagccactgttaagcatgcaa 789
Db    616 TGGCTGAAAGAATCAAAACCTTTTATGCATTAGCTCCAGTCGCCACTGTGAAGTATACAA 675

Qy    790 aaagccccgggaccaaattttttgttgctgccagatatgatgatcaagggattgtttggca 849
Db    676 AAAGCCTTGTAAACAAACTTAGGTTTATTCCTCCAACCATGTTCAAGATTATATTTGGTG 735

Qy    850 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 909
Db    736 ACAAAATATTCTACCCACACAATTTCTTTGATCAATTTCTTGCCACTCAAGTGTGTTCCC 795

Qy    910 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 969
Db    796 GTGAGACACTGAATGTAATTTGCAGCAATGCCTTATTTATCATTTGTGGCTTTGACAGCG 855

Qy    970 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 1029
Db    856 CAAACTTGAACATGAGTCGCTTGGATGTGTACGTGTCACATAATCCGGCAGGAACTTCAG 915

Qy   1030 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 1089
Db    916 TTCAAAACATGCTGCACTGGACCCAGGCTGTTAAATCTGGAAATTTTCAAGCTTTTAATT 975

Qy   1090 ggggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1149
Db    976 GGGGAAGTCCAGCTCAGAATGTAGTGCACTTCAATCAGCCCACACCTCCCTACTACAATG 1035

Qy   1150 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1209
Db   1036 TGACCGCCATGAATGTGCCAATTGCAGTGTGGAGTGGTGGCAATGACTGGTTGGCTGACC 1095
```

```
Qy   1210 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1269
          |  |||| ||  |  |  ||||  |  |  |  | || ||||| ||||| ||| | ||||
Db   1096 CCCAAGATGTTGACCTTTTGCTTCCAAAACTTTCTAATCTCATTTACCACAAGGAGATTC 1155

Qy   1270 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1329
           |  ||    |||| ||||||| |||||||   || ||||||||||     | |||||||
Db   1156 TTCCATATAATCACTTGGATTTTATCTGGGCAATGAATGCTCCTCAAGAAGTTTACAATG 1215

Qy   1330 aaatcatccatctgatgcagcaggaggaga 1359
          |||| ||   | |||||    | ||  | |
Db   1216 AAATTATTTCTATGATGGCAAAAGATAAAA 1245


RESULT   9
A26690/c
LOCUS       A26690       1378 bp    DNA             PAT       05-APR-1995
DEFINITION  Precursor of rabit gastric lipase (comp.).
ACCESSION   A26690
VERSION     A26690.1  GI:905030
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 1378)
  AUTHORS   Benicourt,C., Blanchard,C. and Junien,J.L.
  TITLE     Recombinant gastric lipase from rabbit and pharmaceutical
            compositions
  JOURNAL   Patent: EP 0542629-A 10 19-MAY-1993;
            INSTITUT DE RECHERCHE JOUVEINAL
FEATURES             Location/Qualifiers
     source          1. .1378
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
BASE COUNT      418 a    276 c    287 g    397 t
ORIGIN


  Query Match             23.9%;  Score 489.2;  DB 9;  Length 1378;
  Best Local Similarity   65.0%;  Pred. No. 2.5e-112;
  Matches  722;  Conservative    0;  Mismatches  388;  Indels    0;  Gaps    0;

Qy    250 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 309
          |||| ||||     |||||||||||| | || ||      |    || || ||  ||||
Db   1243 ACCCTGAAGTGAATATGAATATTAGTCAGATGATTTCCTACTGGGGATACCCAAGTGAAA 1184

Qy    310 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 369
          ||||||||||  |  |||||||||||.|||||||||   |||||||| ||||||   ||
Db   1183 AATATGAAGTTGTGACTGAAGATGGTTATATCCTTGAGGTTAACAGAATTCCTTATGGGA 1124

Qy    370 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 429
             |  |   ||| | |||    || || || || ||  ||||||||||  |  |||
Db   1123 AGAAAAATTCAGGGAACAGAGGCCAGAGACCCGTCGTATTTTTGCAGCATGGTTTGCTTG 1064

Qy    430 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 489
          |   ||    |||||||| |||||||||||||||||| ||||||| ||| ||||||||||
Db   1063 CATCAGCCTCAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTTATTCTGGCAG 1004

Qy    490 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 549
          ||||||||| ||  |||||| |||||||||||||| |||||  |||||||  ||| | ||
Db   1003 ATGCTGGTTATGGTGTGTGGCTGGGGAACAGCAGAGGAAATACCTGGTCCAGAAGAAACT 944

Qy    550 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 609
                ||   ||||   | ||| |||||||||||| || | |||||| |||||||   |
Db    943 TATACTATTCACCAGACTCCGTTGAATTCTGGGCTTTTAGCTTTGATGAAATGGCTAAAT 884

Qy    610 ttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaaagatctatt 669
          |||||||||  ||    ||  |||| ||| |  || |||| || |||||||||  |  | |
Db    883 ATGACCTTCCAGCCACAATTGACTTCATTGTAAAGGAAACTGGACAGGAAAAGCTGCACT 824

Qy    670 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 729
          |||| ||  |||| ||||||||||||||| || |||||||| |||||  || |  ||  |||
Db    823 ATGTTGGTCATTCCCAGGGCACCACCATTGGTTTTATTGCCTTTTCTACAAATCCCAAGC 764
```

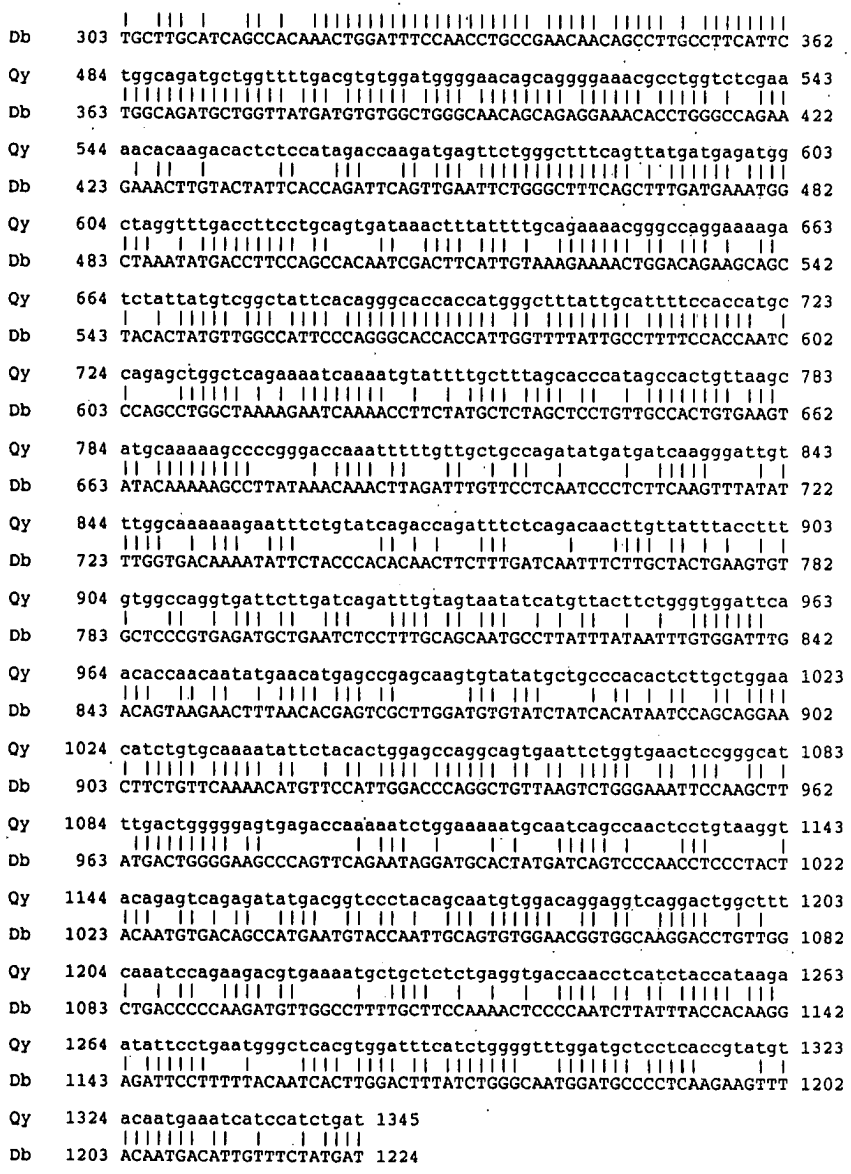
```
Qy    730 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 789
          ||||| | | ||||||||  | || ||| ||||| ||  | |||||||| ||| || |||
Db    763 TGGCTGAAAGAATCAAAACCTTTTATGCATTAGCTCCAGTCGCCACTGTGAAGTATACAA 704

Qy    790 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 849
          ||||||  |  | |||| ||  |||  | ||     |  ||| |||||    | |||||
Db    703 AAAGCCTTGTAAACAAACTTAGGTTTATTCCTCCAACCATGTTCAAGATTATATTTGGTG 644

Qy    850 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 909
          | |||  |||        || |     |||     |  |||| | | |  | |||  ||
Db    643 ACAAAATATTCTACCCACACAATTTCTTTGATCAATTTCTTGCCACTCAAGTGTGTTCCC 584

Qy    910 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 969
          | ||  ||  ||   ||||| || |||  | | ||  |  |  |||| ||  ||| |
Db    583 GTGAGACACTGAATGTAATTTGCAGCAATGCCTTATTTATCATTTGTGGCTTTGACAGCG 524

Qy    970 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 1029
          .||  |||||||||| ||      ||| || |  | || | ||  || |||| || |
Db    523 CAAACTTGAACATGAGTCGCTTGGATGTGTACGTGTCACATAATCCGGCAGGAACTTCAG 464

Qy   1030 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 1089
          | ||||| || || ||||||| |||||| || || |||||  |  | |  || ||| | |
Db    463 TTCAAAACATGCTGCACTGGACCCAGGCTGTTAAATCTGGAAATTTTCAAGCTTTTAATT 404

Qy   1090 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1149
          |||| |||     |  | ||| | |   | | |||||||||| || |||      |||| |
Db    403 GGGGAAGTCCAGCTCAGAATGTAGTGCACTTCAATCAGCCCACACCTCCCTACTACAATG 344

Qy   1150 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1209
          | |  |  |||| || || |  ||| |||||  || ||  | |||||| |  |  | |
Db    343 TGACCGCCATGAATGTGCCAATTGCAGTGTGGAGTGGTGGCAATGACTGGTTGGCTGACC 284

Qy   1210 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1269
          |  |||| ||  |  |  ||||  |  |  |  | || ||||| ||||| ||| | ||||
Db    283 CCCAAGATGTTGACCTTTTGCTTCCAAAACTTTCTAATCTCATTTACCACAAGGAGATTC 224

Qy   1270 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1329
           |  ||    |||| ||||||| |||||||   || ||||||||||     | |||||||
Db    223 TTCCATATAATCACTTGGATTTTATCTGGGCAATGAATGCTCCTCAAGAAGTTTACAATG 164

Qy   1330 aaatcatccatctgatgcagcaggaggaga 1359
          |||| ||   | |||||    | ||  | |
Db    163 AAATTATTTCTATGATGGCAAAAGATAAAA 134


RESULT  10
HSGLR
LOCUS       HSGLR        1365 bp    mRNA            PRI       30-MAR-1995
DEFINITION  Human mRNA for gastric lipase.
ACCESSION   X05997
VERSION     X05997.1  GI:31771
KEYWORDS    gastric lipase; lipase.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1365)
  AUTHORS   Bodmer,M.W., Angal,S., Yarranton,G.T., Harris,T.J., Lyons,A.,
            King,D.J., Pieroni,G., Riviere,C., Verger,R. and Lowe,P.A.
  TITLE     Molecular cloning of a human gastric lipase and expression of the
            enzyme in yeast
  JOURNAL   Biochim. Biophys. Acta 909 (3), 237-244 (1987)
  MEDLINE   87299724
COMMENT     Data kindly reviewed (10-DEC-1987) by LOWE P.A.
FEATURES             Location/Qualifiers
     source          1. .1365
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     sig_peptide     45. .101
                     /product="pot.signal peptide (AA -19 to -1)"
     CDS             45. .1241
                     /codon_start=1
                     /product="gastric lipase precursor"
```

```
                     /protein_id="CAA29413.1"
                     /db_xref="GI:758063"
                     /db_xref="SWISS-PROT:P07098"
                     /translation="MWLLLTMASLISVLGTTHGLFGKLHPGSPEVTMNISQMITYWGY
                     PNEEYEVVTEDGYILEVNRIPYGKKNSGNTGQRPVVFLQHGLLASATNWISNLPNNSL
                     AFILADAGYDVWLGNSRGNTWARRNLYYSPDSVEFWAFSFDEMAKYDLPATIDFIVKK
                     TGQKQLHYVGHSQGTTIGFIAFSTNPSLAKRIKTFYALAPVATVKYTKSLINKLRFVP
                     QSLFKFIFGDKIFYPHNFFDQFLATEVCSREMLNLLCSNALFIICGFDSKNFNTSRLD
                     VYLSHNPAGTSVQNMFHWTQAVKSGKFQAYDWGSPVQNRMHYDQSQPPYYNVTAMNVP
                     IAVWNGGKDLLADPQDVGLLLPKLPNLIYHKEIPFYNHLDFIWAMDAPQEVYNDIVSM
                     ISEDKK"
     CDS             63. .1241
                     /codon_start=1
                     /product="gastric lipase precursor"
                     /protein_id="CAA29414.1"
                     /db_xref="GI:758064"
                     /db_xref="SWISS-PROT:P07098"
                     /translation="MASLISVLGTTHGLFGKLHPGSPEVTMNISQMITYWGYPNEEYE
                     VVTEDGYILEVNRIPYGKKNSGNTGQRPVVFLQHGLLASATNWISNLPNNSLAFILAD
                     AGYDVWLGNSRGNTWARRNLYYSPDSVEFWAFSFDEMAKYDLPATIDFIVKKTGQKQL
                     HYVGHSQGTTIGFIAFSTNPSLAKRIKTFYALAPVATVKYTKSLINKLRFVPQSLFKF
                     IFGDKIFYPHNFFDQFLATEVCSREMLNLLCSNALFIICGFDSKNFNTSRLDVYLSHN
                     PAGTSVQNMFHWTQAVKSGKFQAYDWGSPVQNRMHYDQSQPPYYNVTAMNVPIAVWNG
                     GKDLLADPQDVGLLLPKLPNLIYHKEIPFYNHLDFIWAMDAPQEVYNDIVSMISEDKK
                     "
     sig_peptide     63. .101
                     /product="pot.signal peptide (AA -13 to -1)"
     mat_peptide     102. .1238
                     /product="gastric lipase (AA 1-379)"
BASE COUNT      396 a    284 c    268 g    417 t
ORIGIN
```

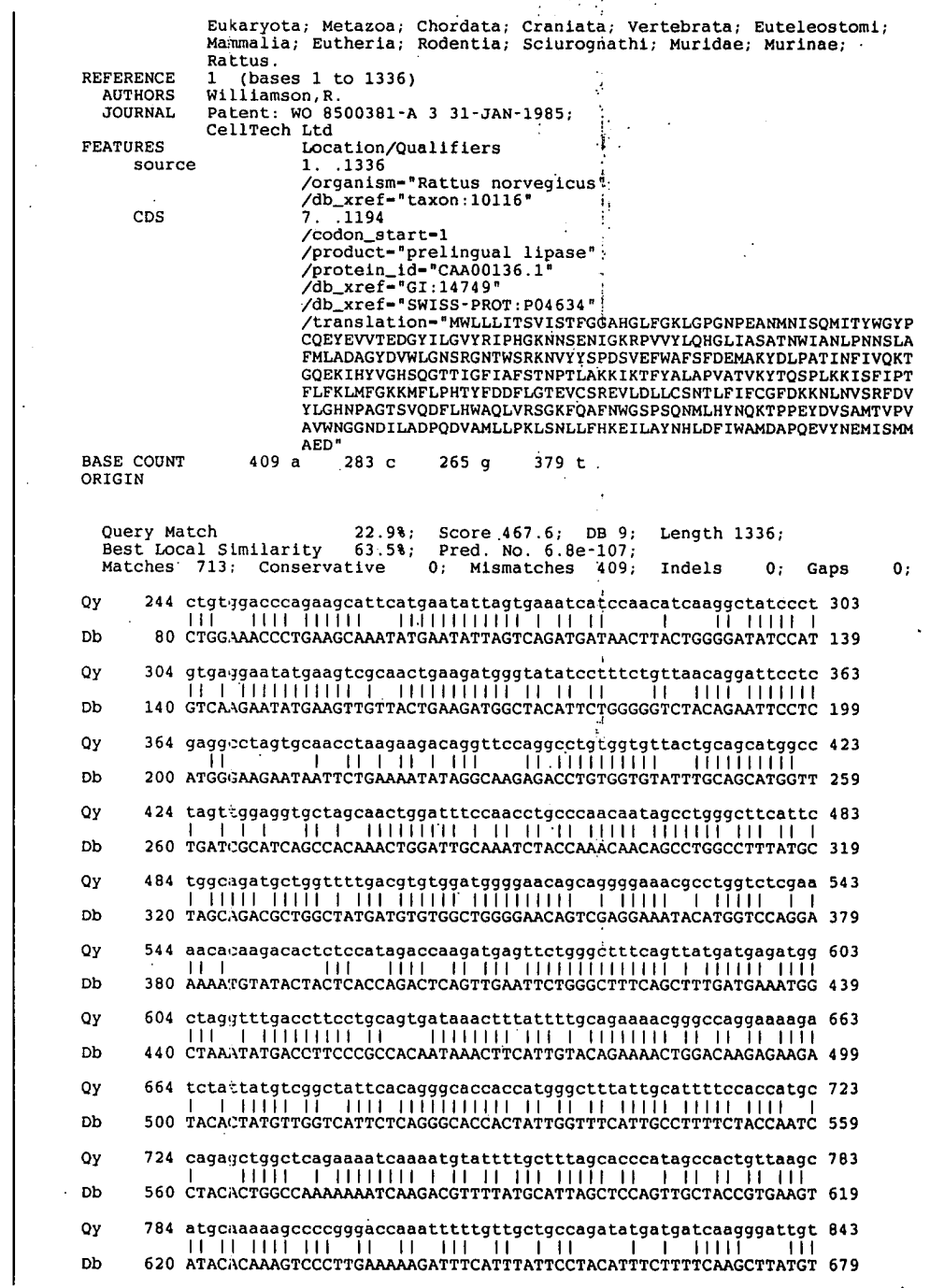
```
  Query Match           23.0%;  Score 470;  DB 93;  Length 1365;
  Best Local Similarity   64.2%;  Pred. No. 1.7e-107;
  Matches  707;  Conservative    0;  Mismatches  395;  Indels    0;  Gaps    0;

Qy     244 ctgtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccct 303
           |||    ||| ||||     ||||| ||||||  || ||     ||  || || ||
Db     121 CTGGAAGCCCTGAAGTGACTATGAACATTAGTCAGATGATTACTTATTGGGGATACCCAA 180

Qy     304 gtgaggaatatgaagtcgcaactgaagatgggtatatccctttctgttaacaggattcctc 363
           ||| |||||||||||  |  ||||||||||| ||||| |||   || || || ||||||
Db     181 ATGAAGAATATGAAGTTGTGACTGAAGATGGTTATATTCTTGAAGTCAATAGAATTCCTT 240

Qy     364 gaggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcc 423
           ||        |  |   ||| |||||    || ||||| |||||  ||||||||||
Db     241 ATGGGAAGAAAAATTCAGGGAATACAGGCCAGAGACCTGTTGTGTTTTTGCAGCATGGTT 300

Qy     424 tagttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattc 483
           |  ||| |   || |  ||||||||||||||||||| ||||| ||||| | ||||||||
Db     301 TGCTTGCATCAGCCACAAACTGGATTTCCAACCTGCCGAACAACAGCCTTGCCTTCATTC 360

Qy     484 tggcagatgctggtttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaa 543
           ||||||||||||||| ||| |||||.|||| |||||||| |||||| ||||| |  |||
Db     361 TGGCAGATGCTGGTTATGATGTGTGGCTGGGCAACAGCAGAGGAAACACCTGGGCCAGAA 420

Qy     544 aacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatgg 603
           | ||  |      ||   |||  || ||| ||||||||||||||| | ||||||  ||||
Db     421 GAAACTTGTACTATTCACCAGATTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGG 480

Qy     604 ctaggtttgaccttcctgcagtgataaactttatttttgcagaaaacgggccaggaaaaga 663
           |||  | |||||||||  ||    ||  |||| ||| |  |||||| || ||| |  ||
Db     481 CTAAATATGACCTTCCAGCCACAATCGACTTCATTGTAAAGAAAACTGGACAGAAGCAGC 540

Qy     664 tctattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgc 723
           |  | ||||| ||| |||| |||||||||||||| || |||||||  ||||||||||  |
Db     541 TACACTATGTTGGCCATTCCCAGGGCACCACCATTGGTTTTATTGCCTTTTCCACCAATC 600

Qy     724 cagagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagc 783
           |    |||||| | | ||||||||  |  | |||| |||| ||  | ||||||||  |||
Db     601 CCAGCCTGGCTAAAAGAATCAAAACCTTCTATGCTCTAGCTCCTGTTGCCACTGTGAAGT 660

Qy     784 atgcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgt 843
           || |||||||||      | |||| ||   ||  | ||  |      |  ||||||    | |
Db     661 ATACAAAAAGCCTTATAAACAAACTTAGATTTGTTCCTCAATCCCTCTTCAAGTTTATAT 720

Qy     844 ttggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccttt 903
           ||||  | |||  |||      || | |   |||     |   |||| || | |  | |
Db     721 TTGGTGACAAAATATTCTACCCACACAACTTCTTTGATCAATTTCTTGCTACTGAAGTGT 780

Qy     904 gtggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattca 963
           |  .||  | ||| ||  |||   |||| || |||  | | ||  |  |  |||||||
Db     781 GCTCCCGTGAGATGCTGAATCTCCTTTGCAGCAATGCCTTATTTATAATTTGTGGATTTG 840

Qy     964 acaccaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaa 1023
           |||  || ||  | |||| ||| ||       ||| |||    | || | ||  || ||||
Db     841 ACAGTAAGAACTTTAACACGAGTCGCTTGGATGTGTATCTATCACATAATCCAGCAGGAA 900

Qy     1024 catctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcat 1083
            | ||||| ||||| ||  | || |||| |||||| || || |||||  || |||  || |
Db     901 CTTCTGTTCAAAACATGTTCCATTGGACCCAGGCTGTTAAGTCTGGGAAATTCCAAGCTT 960

Qy     1084 ttgactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggt 1143
             |||||||||  ||       | |||  |   | |  |||||  |    |||      |
Db     961 ATGACTGGGGAAGCCCAGTTCAGAATAGGATGCACTATGATCAGTCCCAACCTCCCTACT 1020

Qy     1144 acagagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggcttt 1203
            |||  || | ||  |||  || || |  ||| ||||||  || ||  |||||  | |
Db     1021 ACAATGTGACAGCCATGAATGTACCAATTGCAGTGTGGAACGGTGGCAAGGACCTGTTGG 1080

Qy     1204 caaatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataaga 1263
            |    ||  |||| ||      | |||  |  |  |  |||| || || ||||| |||
Db     1081 CTGACCCCCAAGATGTTGGCCTTTTGCTTCCAAAACTCCCCAATCTTATTTACCACAAGG 1140

Qy     1264 atattcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgt 1323
            | ||||||   |    |||| |||| || |||||||   |||||||  |||||     | |
Db     1141 AGATTCCTTTTTACAATCACTTGGACTTTATCTGGGCAATGGATGCCCCTCAAGAAGTTT 1200

Qy     1324 acaatgaaatcatccatctgat 1345
            |||||||  ||  |   | ||||
Db     1201 ACAATGACATTGTTTCTATGAT 1222
```

```
RESULT  11
A01046
LOCUS       A01046       1367 bp    mRNA            PAT       16-MAR-1993
DEFINITION  H.sapiens mRNA for human gastric lipase.
ACCESSION   A01046
VERSION     A01046.1  GI:344241
KEYWORDS    gastric lipase.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1367)
  AUTHORS   .
  TITLE     POLYPEPTIDE AND POLYPEPTIDE COMPOSITION
  JOURNAL   Patent: WO 8601532-A 5 13-MAR-1986;
FEATURES             Location/Qualifiers
     source          1. .1367
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     sig_peptide     47. .103
     CDS             47. .1243
                     /note="gastric lipase"
                     /codon_start=1
                     /product="lipase"
                     /protein_id="CAA00125.1"
                     /db_xref="GI:344242"
                     /db_xref="SWISS-PROT:P07098"
                     /translation="MWLLLTMASLISVLGTTHGLFGKLHPGSPEVTMNISQMITYWGY
                     PNEEYEVVTEDGYILEVNRIPYGKKNSGNTGQRPVVFLQHGLLASATNWISNLPNNSL
                     AFILADAGYDVWLGNSRGNTWARRNLYYSPDSVEFWAFSFDEMAKYDLPATIDFIVKK
                     TGQKQLHYVGHSQGTTIGFIAFSTNPSLAKRIKTFYALAPVATVKYTKSLINKLRFVP
                     QSLFKFIFGDKIFYPHNFFDQFLATEVCSREMLNLLCSNALFIICGFDSKNFNTSRLD
```

```
                 VYLSHNPAGTSVQNMFHWTQAVKSGKFQAYDWGSPVQNRMHYDQSQPPYYNVTAMNVP
                 IAVWNGGKDLLADPQDVGLLLPKLPNLIYHKEIPFYNHLDFIWAMDAPQEVYNDIVSM
                 ISEDKK"
    mat_peptide  104. .1240
                 /note="gastric lipase"
                 /product="lipase"
BASE COUNT      397 a    284 c    269 g    417 t
ORIGIN


  Query Match           23.0%;  Score 470;  DB 9;  Length 1367;
  Best Local Similarity  64.2%;  Pred. No. 1.7e-107;
  Matches  707;  Conservative    0;  Mismatches  395;  Indels    0;  Gaps    0;

Qy    244 ctgtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccct 303
          |||    ||| ||||     ||||| |||||| | || ||     ||   || || ||
Db    123 CTGGAAGCCCTGAAGTGACTATGAACATTAGTCAGATGATTACTTATTGGGGATACCCAA 182

Qy    304 gtgaggaatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctc 363
          ||| ||||||||||| |  |||||||||||| ||||| |||   || || || ||||||
Db    183 ATGAAGAATATGAAGTTGTGACTGAAGATGGTTATATTCTTGAAGTCAATAGAATTCCTT 242

Qy    364 gaggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcc 423
            ||        |  |   ||| |||||     || ||||| |||||  ||||||||||
Db    243 ATGGGAAGAAAAATTCAGGGAATACAGGCCAGAGACCTGTTGTGTTTTTGCAGCATGGTT 302

Qy    424 tagttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattc 483
          |  ||| |   || |  |||||||||||||||||||| ||||| ||||| | ||||||||
Db    303 TGCTTGCATCAGCCACAAACTGGATTTCCAACCTGCCGAACAACAGCCTTGCCTTCATTC 362

Qy    484 tggcagatgctggtttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaa 543
          ||||||||||||||| ||| |||||| |||| |||||||| |||||| ||||| |  |||
Db    363 TGGCAGATGCTGGTTATGATGTGTGGCTGGGCAACAGCAGAGGAAACACCTGGGCCAGAA 422

Qy    544 aacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatgg 603
          | || |      ||   |||   || ||| ||||||||||||||| | |||||| ||||
Db    423 GAAACTTGTACTATTCACCAGATTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGG 482

Qy    604 ctaggtttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaaaga 663
          |||  | |||||||| ||    ||  |||| ||| |  |||||| || ||| |  ||
Db    483 CTAAATATGACCTTCCAGCCACAATCGACTTCATTGTAAAGAAAACTGGACAGAAGCAGC 542

Qy    664 tctattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgc 723
          |  | ||||| ||| |||| |||||||||||||| || |||||||| ||||||||||  |
Db    543 TACACTATGTTGGCCATTCCCAGGGCACCACCATTGGTTTTATTGCCTTTTCCACCAATC 602

Qy    724 cagagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagc 783
          |    |||||| | | |||||||||  |  | |||| |||| ||  |  |||||||| |||
Db    603 CCAGCCTGGCTAAAAGAATCAAAACCTTCTATGCTCTAGCTCCTGTTGCCACTGTGAAGT 662

Qy    784 atgcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgt 843
          || |||||||||     | |||| ||   ||  | ||  |     |  |||||     | |
Db    663 ATACAAAAAGCCTTATAAACAAACTTAGATTTGTTCCTCAATCCCTCTTCAAGTTTATAT 722

Qy    844 ttggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttt 903
          ||||  | |||  |||       || | |   |||      |   |||| || | |  | |
Db    723 TTGGTGACAAAATATTCTACCCACACAACTTCTTTGATCAATTTCTTGCTACTGAAGTGT 782

Qy    904 gtggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattca 963
          |   ||  | ||| ||  |||   |||| || |||  | | ||  |  | ||||||||
Db    783 GCTCCCGTGAGATGCTGAATCTCCTTTGCAGCAATGCCTTATTTATAATTTGTGGATTTG 842

Qy    964 acaccaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaa 1023
          |||  || ||  | |||| || ||      ||| |||     | || | ||  || ||||
Db    843 ACAGTAAGAACTTTAACACGAGTCGCTTGGATGTGTATCTATCACATAATCCAGCAGGAA 902

Qy   1024 catctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcat 1083
          | |||||  |||| ||  | || |||| |||||| || || |||||  || |||  || |
Db    903 CTTCTGTTCAAAACATGTTCCATTGGACCCAGGCTGTTAAGTCTGGGAAATTCCAAGCTT 962

Qy   1084 ttgactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggt 1143
          |||||||||  ||        | |||  |    | |   |||||  |    |||      |
Db    963 ATGACTGGGGAAGCCCAGTTCAGAATAGGATGCACTATGATCAGTCCCAACCTCCCTACT 1022

Qy   1144 acagagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggcttt 1203
          |||  || | ||  ||||  || || |  ||| |||||| || ||  |||||  | |
Db   1023 ACAATGTGACAGCCATGAATGTACCAATTGCAGTGTGGAACGGTGGCAAGGACCTGTTGG 1082

Qy   1204 caaatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataaga 1263
          |  | ||  |||| ||      |  ||||  |  |  |  |||| || || ||||| |||
Db   1083 CTGACCCCCAAGATGTTGGCCTTTTGCTTCCAAAACTCCCCAATCTTATTTACCACAAGG 1142

Qy   1264 atattcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgt 1323
          | |||||||   |    |||| |||| || |||||||   ||||||| |||||     | |
Db   1143 AGATTCCTTTTTACAATCACTTGGACTTTATCTGGGCAATGGATGCCCCTCAAGAAGTTT 1202

Qy   1324 acaatgaaatcatccatctgat 1345
          ||||||| ||  |   | ||||
Db   1203 ACAATGACATTGTTTCTATGAT 1224


RESULT  12
A12714
LOCUS       A12714       1367 bp    DNA             PAT       09-AUG-1994
DEFINITION  Pregastric lipase.
ACCESSION   A12714
VERSION     A12714.1  GI:579560
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1367)
  AUTHORS   .
  TITLE     PRECURSOR POLYPEPTIDE, DNA SEQUENCE CODING THEREFOR, VECTORS, HOST
            ORGANISMS, AND PROCESSES INVOLVING SAME
  JOURNAL   Patent: WO 8603778-A 8 03-JUL-1986;.
FEATURES             Location/Qualifiers
     source          1. .1367
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             47. .1243
                     /codon_start=1
                     /product="pregastric lipase"
                     /protein_id="CAA01053.1"
                     /db_xref="GI:579561"
                     /db_xref="SWISS-PROT:P07098"
                     /translation="MWLLLTMASLISVLGTTHGLFGKLHPGSPEVTMNISQMITYWGY
                     PNEEYEVVTEDGYILEVNRIPYGKKNSGNTGQRPVVFLQHGLLASATNWISNLPNNSL
                     AFILADAGYDVWLGNSRGNTWARRNLYYSPDSVEFWAFSFDEMAKYDLPATIDFIVKK
                     TGQKQLHYVGHSQGTTIGFIAFSTNPSLAKRIKTFYALAPVATVKYTKSLINKLRFVP
                     QSLFKFIFGDKIFYPHNFFDQFLATEVCSREMLNLLCSNALFIICGFDSKNFNTSRLD
                     VYLSHNPAGTSVQNMFHWTQAVKSGKFQAYDWGSPVQNRMHYDQSQPPYYNVTAMNVP
                     IAVWNGGKDLLADPQDVGLLLPKLPNLIYHKEIPFYNHLDFIWAMDAPQEVYNDIVSM
                     ISEDKK"
BASE COUNT      397 a    284 c    269 g    417 t
ORIGIN


  Query Match           23.0%;  Score 470;  DB 9;  Length 1367;
  Best Local Similarity  64.2%;  Pred. No. 1.7e-107;
  Matches  707;  Conservative    0;  Mismatches  395;  Indels    0;  Gaps    0;

Qy    244 ctgtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccct 303
          |||    ||| ||||     ||||| |||||| | || ||     ||   || || ||
Db    123 CTGGAAGCCCTGAAGTGACTATGAACATTAGTCAGATGATTACTTATTGGGGATACCCAA 182

Qy    304 gtgaggaatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctc 363
          ||| ||||||||||| |  |||||||||||| ||||| |||   || || || ||||||
Db    183 ATGAAGAATATGAAGTTGTGACTGAAGATGGTTATATTCTTGAAGTCAATAGAATTCCTT 242

Qy    364 gaggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcc 423
            ||        |  |   ||| |||||     || ||||| |||||  ||||||||||
Db    243 ATGGGAAGAAAAATTCAGGGAATACAGGCCAGAGACCTGTTGTGTTTTTGCAGCATGGTT 302

Qy    424 tagttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattc 483
```



RESULT 13
A01157
LOCUS A01157 1336 bp mRNA PAT 29-JAN-1993
DEFINITION R.norvegicus mRNA for prelingual lipase protein.
ACCESSION A01157
VERSION A01157.1 GI:14748
KEYWORDS lipase.
SOURCE Norway rat.
ORGANISM Rattus norvegicus



Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 1336)
AUTHORS Williamson,R.
JOURNAL Patent: WO 8500381-A 3 31-JAN-1985;
CellTech Ltd
FEATURES Location/Qualifiers
source 1..1336
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
CDS 7..1194
/codon_start=1
/product="prelingual lipase"
/protein_id="CAA00136.1"
/db_xref="GI:14749"
/db_xref="SWISS-PROT:P04634"
/translation="MWLLLITSVISTFGCAHGLFGKLGPGNPEANMNISQMITYWGYP
CQEYEVVTEDGYILGVYRIPHGKNNSENIGKRPVVYLQHGLIASATNWIANLPNNSLA
FMLADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATINFIVQKT
GQEKIHYVGHSQGTTIGFIAFSTNPTLAKKIKTFYALAPVATVKYTQSPLKKISFIPT
FLFKLMFGKKMFLPHTYFDDFLGTEVCSREVLDLLCSNTLFIFCGFDKKNLNVSRFDV
YLGHNPAGTSVQDFLHWAQLVRSGKFQAFNWGSPSQNMLHYNQKTPPEYDVSAMTVPV
AVWNGGNDILADPQDVAMLLPKLSNLLFHKEILAYNHLDFIWAMDAPQEVYNEMISMM
AED"
BASE COUNT 409 a 283 c 265 g 379 t
ORIGIN

Query Match 22.9%; Score 467.6; DB 9; Length 1336;
Best Local Similarity 63.5%; Pred. No. 6.8e-107;
Matches 713; Conservative 0; Mismatches 409; Indels 0; Gaps 0;

```
Qy     844 ttggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccttt 903
           ||||||| |||   || |||  || |||   |||        |||| ||   |  | |
Db     680 TTGGCAAGAAAATGTTCCTGCCCCACACCTACTTTGATGACTTTCTTGGTACCGAAGTGT 739

Qy     904 gtggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattca 963
           |    | || | |||| ||||   | || || || |   | ||  |  |  |||||||
Db     740 GCTCACGGGAGGTTCTAGATCTTCTCTGCAGCAACACTTTATTCATCTTCTGTGGATTTG 799

Qy     964 acaccaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaa 1023
           |||  || ||  | ||  |||| ||      ||| |||   | || | ||  || || |
Db     800 ACAAGAAAAACTTAAATGTGAGTCGTTTTGATGTGTATCTAGGGCATAATCCAGCAGGGA 859

Qy    1024 catctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcat 1083
           |||||| ||| |  |||| ||||||   |||   || |  |||||  || | |  || |
Db     860 CATCTGTTCAAGACTTTCTCCACTGGGCACAGCTTGTTAGATCTGGGAAATTTCAAGCCT 919

Qy    1084 ttgactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggt 1143
           |  ||||||| ||     || | ||  ||  | | | ||| |||  |||.|||      |
Db     920 TCAACTGGGGAAGCCCATCCCAGAACATGTTACACTACAACCAGAAAACGCCTCCTGAAT 979

Qy    1144 acagagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggcttt 1203
           |    ||  ||. ||||| || ||    ||| ||||||  || ||  | |||   ||
Db     980 ATGATGTGTCAGCCATGACTGTGCCAGTTGCAGTGTGGAACGGTGGCAATGACATCCTGG 1039

Qy    1204 caaatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataaga 1263
           |  ||||  |||| ||    ||||||||  |  |  | ||||||||  | |  |||||||
Db     1040 CTGATCCCCAAGATGTCGCCATGCTGCTTCCCAAACTCTCCAACCTCCTGTTCCATAAGG 1099

Qy    1264 atattcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgt 1323
           | |||| ||  |    |||| |||| ||||||||||   |||||| ||||     | |
Db     1100 AGATTCTTGCCTACAATCACCTGGACTTCATCTGGGCAATGGATGCCCCTCAAGAGGTTT 1159

Qy    1324 acaatgaaatcatccatctgatgcagcaggaggagaccaacc 1365
           ||||||| || ||     |||||    | ||  | |  || |
Db     1160 ACAATGAGATGATTTCCATGATGGCAGAAGACTAAAAGAATC 1201


RESULT  14
RNLIP
LOCUS       RNLIP        1355 bp    mRNA            ROD       30-MAR-1995
DEFINITION  Rat mRNA for lingual lipase.
ACCESSION   X02309
VERSION     X02309.1  GI:56595
KEYWORDS    glycoprotein; lipase.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1355)
  AUTHORS   Docherty,A.J., Bodmer,M.W., Angal,S., Verger,R., Riviere,C.,
            Lowe,P.A., Lyons,A., Emtage,J.S. and Harris,T.J.
  TITLE     Molecular cloning and nucleotide sequence of rat lingual lipase
            cDNA
  JOURNAL   Nucleic Acids Res. 13 (6), 1891-1903 (1985)
  MEDLINE   85215587
COMMENT     Data kindly reviewed (12-FEB-1986) by A.J.P. Docherty.
FEATURES             Location/Qualifiers
     source          1. .1355
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
     sig_peptide     49. .102
                     /product="put. signal peptide (aa -18 to -1)"
     CDS             49. .1236
                     /note="lipase precursor"
                     /codon_start=1
                     /protein_id="CAA26179.1"
                     /db_xref="GI:56596"
                     /db_xref="SWISS-PROT:P04634"
                     /translation="MWLLLITSVISTFGGAHGLFGKLGPGNPEANMNISQMITYWGYP
                     CQEYEVVTEDGYILGVYRIPHGKNNSENIGKRPVVYLQHGLIASATNWIANLPNNSLA
                     FMLADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATINFIVQKT
                     GQEKIHYVGHSQGTTIGFIAFSTNPTLAKKIKTFYALAPVATVKYTQSPLKKISFIPT
                     FLFKLMFGKKMFLPHTYFDDFLGTEVCSREVLDLLCSNTLFIFCGFDKKNLNVSRFDV
                     YLGHNPAGTSVQDFLHWAQLVRSGKFQAFNWGSPSQNMLHYNQKTPPEYDVSAMTVPV
                     AVWNGGNDILADPQDVAMLLPKLSNLLFHKEILAYNHLDFIWAMDAPQEVYNEMISMM
                     AED"
     mat_peptide     103. .1233
                     /product="lipase (aa 1-377)"
     misc_feature    505. .507
                     /note="aa 153 pot. lipid binding site"
     misc_feature    1355
                     /note="polyA site"
BASE COUNT      400 a    291 c    276 g    387 t      1 others
ORIGIN


  Query Match           22.9%;  Score 467.6;  DB 95;  Length 1355;
  Best Local Similarity  63.5%;  Pred. No. 6.8e-107;
  Matches  713;  Conservative    0;  Mismatches  409;  Indels    0;  Gaps    0;

Qy     244 ctgtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccct 303
           |||   |||| ||||||   ||||||||||||| | || ||     |    || ||||| |
Db     122 CTGGAAACCCTGAAGCAAATATGAATATTAGTCAGATGATAACTTACTGGGGATATCCAT 181

Qy     304 gtgaggaatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctc 363
           || | |||||||||||| |  |||||||||||| || || ||    ||  |||| |||||||
Db     182 GTCAAGAATATGAAGTTGTTACTGAAGATGGCTACATTCTGGGGGTCTACAGAATTCCTC 241

Qy     364 gaggccctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcc 423
           ||        |  || | || | |||    || ||||||||||   ||||||||||
Db     242 ATGCGAAGAATAATTCTGAAAATATAGGCAAGAGACCTGTGGTGTATTTGCAGCATGGTT 301

Qy     424 tagttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattc 483
           |  | | |   || |  |||||||||| | || || || ||||| ||||||| ||| || |
Db     302 TGATCGCATCAGCCACAAACTGGATTGCAAATCTACCAAACAACAGCCTGGCCTTTATGC 361

Qy     484 tggcagatgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaa 543
           | ||||| ||||| | ||| |||||| ||||||||||  | |||||  | ||||| | |
Db     362 TAGCAGACGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGAAATACATGGTCCAGGA 421

Qy     544 aacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagatgg 603
           || |          |||   ||||  || ||| |||||||||||||| | |||||| ||||
Db     422 AAAATGTATACTACTCACCAGACTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGG 481

Qy     604 ctaggtttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaaga 663
           |||  | |||||||||  ||    |||||||| ||| | |||||||| || || || ||||
Db     482 CTAAATATGACCTTCCCGCCACAATAAACTTCATTGTACAGAAAACTGGACAAGAGAAGA 541

Qy     664 tctattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgc 723
           |  | ||||| ||  |||| |||||||||||| || || || ||||| ||||| ||||  |
Db     542 TACACTATGTTGGTCATTCTCAGGGCACCACTATTGGTTTCATTGCCTTTTCTACCAATC 601

Qy     724 cagagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagc 783
           |    |||||  | |||||||| | || || ||| |||| ||  | || || || |||
Db     602 CTACACTGGCCAAAAAAATCAAGACGTTTTATGCATTAGCTCCAGTTGCTACCGTGAAGT 661

Qy     784 atgcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgt 843
           || || |||| |||  ||  ||  |||  ||  | ||    |  |  |||||    |||
Db     662 ATACACAAAGTCCCTTGAAAAAGATTTCATTTATTCCTACATTTCTTTTCAAGCTTATGT 721

Qy     844 ttggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccttt 903
           |||||||  |||    || |||   || |||    |||         |||| ||   |  | |
Db     722 TTGGCAAGAAAATGTTCCTGCCCCACACCTACTTTGATGACTTTCTTGGTACCGAAGTGT 781

Qy     904 gtggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattca 963
           |    | || | |||| ||||   | || || || |   | ||  |  |  |||||||
Db     782 GCTCACGGGAGGTTCTAGATCTTCTCTGCAGCAACACTTTATTCATCTTCTGTGGATTTG 841

Qy     964 acaccaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaa 1023
           |||  || ||  | ||  |||| ||      ||| ||| ; |  || | ||  || || |
Db     842 ACAAGAAAAACTTAAATGTGAGTCGTTTTGATGTGTATCTAGGGCATAATCCAGCAGGGA 901

Qy    1024 catctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcat 1083
           |||||| ||| |  |||| ||||||   |||   || |  |||||  || | |  || |
Db     902 CATCTGTTCAAGACTTTCTCCACTGGGCACAGCTTGTTAGATCTGGGAAATTTCAAGCCT 961
```

```
Qy   1084 ttgactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggt 1143
          |  ||||||| ||     || | ||  ||  | | | ||| |||  ||| |||      |
Db    962 TCAACTGGGGAAGCCCATCCCAGAACATGTTACACTACAACCAGAAAACGCCTCCTGAAT 1021

Qy   1144 acagagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggcttt 1203
          |    ||   ||  ||||| || ||    ||| ||||||  || ||  | |||   ||
Db   1022 ATGATGTGTCAGCCATGACTGTGCCAGTTGCAGTGTGGAACGGTGGCAATGACATCCTGG 1081

Qy   1204 caaatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataaga 1263
          |  ||||  |||| ||    |||||||| |  |  |  |||||||| | | |||||||
Db   1082 CTGATCCCCAAGATGTCGCCATGCTGCTTCCCAAACTCTCCAACCTCCTGTTCCATAAGG 1141

Qy   1264 atattcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgt 1323
          | |||| ||  |    |||| |||| ||||||||||   ||||||| |||||     | |
Db   1142 AGATTCTTGCCTACAATCACCTGGACTTCATCTGGGCAATGGATGCCCCTCAAGAGGTTT 1201

Qy   1324 acaatgaaatcatccatctgatgcagcaggaggagaccaacc 1365
          |||||||  || ||     |||||    | ||  | |  || |
Db   1202 ACAATGAGATGATTTCCATGATGGCAGAAGACTAAAAGAATC 1243


RESULT  15
MMLYACLY
LOCUS       MMLYACLY     2358 bp    mRNA            ROD       24-FEB-1999
DEFINITION  M.musculus (C57 Black/6X CBA) LAL mRNA for lysosomal acid lipase.
ACCESSION   Z31689
VERSION     Z31689.1  GI:4456670
KEYWORDS    LAL; lipase; lysosomal acid lipase.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2358)
  AUTHORS   Du,H., Witte,D.P. and Grabowski,G.A.
  TITLE     Tissue and cellular specific expression of murine lysosomal acid
            lipase mRNA and protein
  JOURNAL   J. Lipid Res. 37 (5), 937-949 (1996)
  MEDLINE   96363957
REFERENCE   2  (bases 1 to 2358)
  AUTHORS   Du,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-APR-1994) Hong Du, Division of Human Genetics,
            Children's Hospital Medical Center, 3333 Bernet Street, Cincinnati,
            OH, 45229-3039, USA
FEATURES             Location/Qualifiers
     source          1. .2358
                     /organism="Mus musculus"
                     /strain="C57 Black/6X CBA"
                     /db_xref="taxon:10090"
                     /sex="Female"
                     /clone="mLAL #49"
                     /tissue_type="liver"
                     /clone_lib="Mouse liver cDNA library /Lamda-ZAP"
     gene            19. .1212
                     /gene="LAL"
     CDS             19. .1212
                     /gene="LAL"
                     /citation=[1]
                     /codon_start=1
                     /product="lysosomal acid lipase"
                     /protein_id="CAA83494.1"
                     /db_xref="GI:4456671"
                     /translation="MQLQGLVFVFTIGILLSRVPTGTVSAVDPEVNMNVTEIIMRWGY
                     PGEEHSVLTGDGYILSIHRIPRGWKNHFGKGPRPVVYLQHGLLADSSNWVTNIDNSSL
                     GFLLADRGFDVWMGNSRGNTWSLKHKTLSVSQDEFWAFSFDEMAKYDLPASINYILNK
                     TGQEQIYYVGHSQGCTIGFIAFSQMPELAKKIKMFLVLAPVLSLNFASGPLLQLGRLP
                     DPLLKDMFGQKQFLPQSAMLKWLSIHVCTHVIMKELCANVFFLLCGFNEKNLNMSRVD
                     VYTTHCPAELLVQNMLHWGQVFKYRKLQAFDWGSSEKNYFHYNQSFPPSYNIKNMRLP
                     TALWSGGRDWLADINDITILLTQIPKLVYHKNIPEWDHLDFIWGLDAPWKLYDEIISL
                     MKKYQ"
     sig_peptide     19. .75
                     /gene="LAL"
     mat_peptide     76. .1209
                     /gene="LAL"
                     /product="lysosomal acid lipase"
     polyA_signal    2246. .2251
     polyA_site      2358
BASE COUNT      604 a    549 c    524 g    681 t
ORIGIN


  Query Match             22.6%;  Score 461;  DB 94;  Length 2358;
  Best Local Similarity   63.3%;  Pred. No. 3.3e-105;
  Matches  707;  Conservative    0;  Mismatches  410;  Indels    0;  Gaps    0;

Qy    242 agctgtggacccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatcc 301
          |||||||||| || ||||    |||||||| | |  || || |||    |     || |||||
Db     93 AGCTGTGGATCCTGAAGTCAACATGAATGTGACCGAGATAATCATGCGCTGGGGATATCC 152

Qy    302 ctgtgaggaatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcc 361
            |||||||| |    |||   || | |||||| ||||| ||   | || || | || ||
Db    153 TGGTGAGGAACACTCGGTCCTGACAGGAGATGGCTATATTCTGAGTATTCACCGAATCCC 212

Qy    362 tcgaggcctagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatgg 421
          ||| ||      | | | |     | |||| |||| || || ||||  || ||||| ||
Db    213 TCGTGGGTGGAAGAACCATTTTGGCAAAGGTCCCAGACCAGTTGTGTATCTTCAGCACGG 272

Qy    422 cctagttggaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcat 481
          | |  | | || | |||| ||||||  |  | ||| |   |||||  ||| |||||||| |
Db    273 CTTGCTGGCAGATTCTAGTAACTGGGTCACAAACATTGACAACAGCAGCTTGGGCTTCCT 332

Qy    482 tctggcagatgctggtttttgacgtgtggatggggaacagcaggggaaacgcctggtctcg 541
          |||||  |||  ||| ||||| ||||||||||| |||||||| |||||| |||||||||
Db    333 CCTGGCGGATCGTGGCTTTGATGTGTGGATGGGAAACAGCAGAGGAAACACCTGGTCTCT 392

Qy    542 aaaacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgagat 601
            || |||||||| || ||  |    || |||||||||||||| ||||||| |||||||||
Db    393 GAAGCACAAGACTCTTTCAGTTTCTCAGGATGAGTTCTGGGCCTTCAGTTTTGATGAGAT 452

Qy    602 ggctaggtttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaa 661
          ||| |  | |||||| |||||   || ||||  |||||| | ||||| ||||| ||| |
Db    453 GGCAAAATATGACCTACCTGCTTCAATTAACTACATTTTGAATAAAACTGGCCAAGAACA 512

Qy    662 gatctattatgtcggctattcacagggcaccaccatgggctttattgcattttccaccat 721
          ||||||||||| ||  |||| || |||  |||||| || || || || |||||    ||
Db    513 AATCTATTATGTGGGTCATTCTCAAGGCTGCACCATAGGTTTCATAGCTTTTTCACAGAT 572

Qy    722 gccagagctggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaa 781
          ||| ||| ||||  |||| || ||||||| ||| |  || || ||| |     || | ||
Db    573 GCCTGAGTTGGCAAAGAAGATTAAAATGTTTTTAGTCTTGGCTCCCGTGTTGTCTCTCAA 632

Qy    782 gcatgcaaaaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggatt 841
            |||   | ||||      ||||       |||||||     |  ||||||   |
Db    633 TTTTGCTTCAGGCCCGCTACTTCAATTGGGACGCTTGCCAGACCCTCTTCTCAAGGACAT 692

Qy    842 gtttggcaaaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacct 901
          ||||||  | ||  |||||||   |||||      | || |     ||    ||| || |
Db    693 GTTTGGACAGAAGCAATTTCTCCCTCAGAGTGCGATGCTGAAGTGGCTGAGCATTCACGT 752

Qy    902 ttgtggccaggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggatt 961
          |||   ||| || ||  |  |  || ||||    ||  || | || || ||  |||| ||
Db    753 TTGCACCCACGTCATCATGAAGGAGCTTTGCGCAAACGTCTTCTTCCTGCTATGTGGCTT 812

Qy    962 caacaccaacaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctgg 1021
          |||    || ||| | || |||    |||   ||| ||| |  | |||   | ||| |
Db    813 CAATGAGAAGAATTTAAATATGTCTAGAGTGGATGTGTATACAACACACTGCCCTGCGGA 872

Qy   1022 aacatctgtgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggc 1081
          |      |||||||| ||  |||||||| | ||||   | || |   || | || |  ||
Db    873 ACTTCTCGTGCAAAACATGTTACACTGGGGTCAGGTTTTTAAATACCGTAAGCTTCAAGC 932

Qy   1082 atttgactgggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaag 1141
          |||||||||||| ||       || |||       | | ||| |||     ||||
Db    933 TTTTGACTGGGGAAGCAGTGAAAAGAATTATTTTCATTACAACCAGAGTTTTCCTCCCTC 992
```

```
Qy   1142 gtacagagtcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggct 1201
          |||||    | | | | |||    | || || ||  ||||||  || || | |||||||||
Db    993 GTACAACATAAAAAACATGCGTCTACCCACTGCTTTGTGGAGCGGGGGCCGGGACTGGCT 1052

Qy   1202 ttcaaatccagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataa 1261
            || |     | ||| | |  ||  | ||  || |  |  |||| | | || || ||
Db   1053 GGCAGACATCAATGACATCACTATCTTACTGACTCAAATCCCCAAGTTGGTGTATCACAA 1112

Qy   1262 gaatattcctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtat 1321
          ||| |||||||| ||||  ||  ||||||| ||||||||||||||||| ||        |
Db   1113 GAACATTCCTGAGTGGGACCATCTGGATTTTATCTGGGGTTTGGATGCCCCCTGGAAACT 1172

Qy   1322 gtacaatgaaatcatccatctgatgcagcaggaggag 1358
          |||  ||||||| |||  ||| ||| || |  |  ||
Db   1173 GTATGATGAAATAATCAGTCTAATGAAGAAATACCAG 1209
```

Search completed: August 24, 2001, 22:44:11
Job time: 9133 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: August 24, 2001, 20:38:04 ; Search time 97.86 Seconds
(without alignments)
3954.138 Million cell updates/sec

Title: US-09-333-159-45
Perfect score: 2044
Sequence: 1 gtcgacccacgcgtccgggg.....aaaaaaaaaagggcggccgc 2044

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	444.2	21.7	1137	1	US-08-227-108-2	Sequence 2, Appli
	2	444.2	21.7	1137	2	US-09-073-674-2	Sequence 2, Appli
	3	444.2	21.7	1140	1	US-08-227-108-4	Sequence 4, Appli
	4	444.2	21.7	1140	2	US-09-073-674-4	Sequence 4, Appli
	5	444.2	21.7	1146	1	US-08-227-108-6	Sequence 6, Appli
	6	444.2	21.7	1146	2	US-09-073-674-6	Sequence 6, Appli
	7	444.2	21.7	1528	1	US-08-227-108-1	Sequence 1, Appli
	8	444.2	21.7	1528	2	US-09-073-674-1	Sequence 1, Appli
	9	48.2	2.4	178	1	US-08-751-782-2	Sequence 2, Appli
	10	48.2	2.4	178	2	US-08-925-171-2	Sequence 2, Appli
	11	46.4	2.3	43676	3	US-09-356-952-12	Sequence 12, Appl
c	12	42	2.1	8041	1	US-08-765-081-1	Sequence 1, Appli
c	13	42	2.1	8041	3	US-09-098-082-1	Sequence 1, Appli
c	14	42	2.1	8041	5	PCT-US95-06994-1	Sequence 1, Appli
	15	40.4	2.0	1140	2	US-08-698-805-5	Sequence 5, Appli
	16	38.6	1.9	2815	1	US-08-230-491A-1	Sequence 1, Appli
	17	38.6	1.9	2815	1	US-08-619-280A-1	Sequence 1, Appli
	18	38.6	1.9	2815	2	US-08-940-391-1	Sequence 1, Appli
	19	38.4	1.9	2798	4	US-09-318-448-34	Sequence 34, Appl
	20	36.8	1.8	1248	2	US-08-897-340-5	Sequence 5, Appli
	21	36.8	1.8	1248	3	US-09-252-329-5	Sequence 5, Appli
	22	36.8	1.8	1406	3	US-08-913-842-6	Sequence 6, Appli
	23	36.6	1.8	3271	2	US-08-852-806-1	Sequence 1, Appli
	24	36.6	1.8	3271	3	US-09-163-669-1	Sequence 1, Appli
	25	36.6	1.8	3282	1	US-08-276-852-154	Sequence 154, App
c	26	36.6	1.8	3282	1	US-08-276-852-169	Sequence 169, App
	27	36.6	1.8	3282	1	US-08-899-575-154	Sequence 154, App
c	28	36.6	1.8	3282	1	US-08-899-575-169	Sequence 169, App
	29	36.6	1.8	3282	1	US-08-899-575-154	Sequence 154, App
c	30	36.6	1.8	3282	1	US-08-899-575-169	Sequence 169, App
	31	36.6	1.8	3282	5	PCT-US95-08743-154	Sequence 154, App
c	32	36.6	1.8	3282	5	PCT-US95-08743-169	Sequence 169, App
	33	36.6	1.8	13254	1	US-08-276-852-156	Sequence 156, App
c	34	36.6	1.8	13254	1	US-08-276-852-170	Sequence 170, App
	35	36.6	1.8	13254	1	US-08-899-575-156	Sequence 156, App
c	36	36.6	1.8	13254	1	US-08-899-575-170	Sequence 170, App
	37	36.6	1.8	13254	1	US-08-899-575-156	Sequence 156, App
c	38	36.6	1.8	13254	1	US-08-899-575-170	Sequence 170, App
	39	36.6	1.8	13254	5	PCT-US95-08743-156	Sequence 156, App
c	40	36.6	1.8	13254	5	PCT-US95-08743-170	Sequence 170, App
c	41	36.4	1.8	234	3	US-08-651-472-89	Sequence 89, Appl
c	42	35.8	1.8	6027	2	US-08-968-542C-1	Sequence 1, Appli
c	43	35.2	1.7	37895	1	US-08-375-709-1	Sequence 1, Appli
c	44	35.2	1.7	37895	1	US-08-752-929-1	Sequence 1, Appli
c	45	35.2	1.7	37895	3	US-09-090-793-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT 1
US-08-227-108-2
; Sequence 2, Application US/08227108
; Patent No. 5807726
; GENERAL INFORMATION:
;   APPLICANT: Blanchard, Claire
;   APPLICANT: Benicourt, Claude
;   APPLICANT: Junien, Jean-Louis
;   TITLE OF INVENTION: Recombinant Dog Gastric Lipase
;   NUMBER OF SEQUENCES: 21
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Pennie & Edmonds
;     STREET: 1155 Avenue of the Americas
;     CITY: New York
;     STATE: New York
;     COUNTRY: U.S.A.
;     ZIP: 10036
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/227,108
;     FILING DATE: 03-APR-1994
;     CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Fanucci, Allan A.
;     REGISTRATION NUMBER: 30,256
;     REFERENCE/DOCKET NUMBER: 7620-033
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 212 790-9090
;     TELEFAX: 212 869-8864/9741
;     TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1137 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: double
;     TOPOLOGY: linear
;   MOLECULE TYPE: cDNA to mRNA
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 1..1137
US-08-227-108-2

  Query Match 21.7%; Score 444.2; DB 1; Length 1137;
  Best Local Similarity 62.8%; Pred. No. 1.1e-117;
  Matches 689; Conservative 0; Mismatches 408; Indels 0; Gaps 0;
```

```
Qy    250 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 309
          |||| ||||    |||||||||| ||| | || |||     |    || || ||   |||||
Db     26 ACCCTGAAGTGACCATGAATATAAGTCAGATGATCACCTACTGGGGATACCCAGCTGAGG 85

Qy    310 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 369
          |||||||||| |  || ||||| || |||||||||     |  |||| ||||||   ||
Db     86 AATATGAAGTTGTGACCGAAGACGGTTATATCCTTGGGATCGACAGAATTCCTTATGGGA 145

Qy    370 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 429
             |  | |||| | |||    || ||||| |  ||  |||| || ||  |  | |
Db    146 GGAAAAATTCAGAGAATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCG 205

Qy    430 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 489
          |   || |  |||||||| |||||||||||||||| ||||||| |||||  | |  |
Db    206 CATCAGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCCTGGCCG 265

Qy    490 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 549
          | || || |  |||||||| |||||||||||||||| ||| ||||| |  | |   |
Db    266 ACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACCTGGGCCAGGAGGAATC 325

Qy    550 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 609
            |   |||    |||   |  || |||||||||||||| | ||| ||||||||||  |
Db    326 TGTACTACTCGCCCGACTCCGTCGAATTCTGGGCTTTCAGCTTTGACGAGATGGCTAAAT 385

Qy    610 ttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 669
          |||||||||  ||    ||  |||| || ||| |||||||||| ||||| ||| |  | |
Db    386 ATGACCTTCCCGCCACCATTGACTTCATCTTGAAGAAAACGGGACAGGACAAGCTACACT 445

Qy    670 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 729
          | || ||| |||| |||||||||||||| || || || || |||||||||| ||  |||
Db    446 ACGTTGGCCATTCCCAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGC 505

Qy    730 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 789
          ||||  |    |||||||  |  | ||| ||||| ||| | ||||| || ||| |  |
Db    506 TGGCGAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGTACACCG 565

Qy    790 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 849
          ||| ||     | |||| |  || |  | ||     |  | |||||    | ||||| |
Db    566 AAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTCAAGCTTATATTTGGAA 625

Qy    850 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 909
          | |||  |||       ||   |   |||      |   || |  |  |  | ||   ||
Db    626 ACAAAATATTCTACCCACACCACTTCTTTGATCAATTTCTCGCCACCGAGGTATGCTCCC 685

Qy    910 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 969
            | ||   | ||||   | || || ||   | ||||  |  |  |||||||  |||| |
Db    686 GCGAGACGGTGGATCTCCTCTGCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTA 745

Qy    970 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 1029
            ||  ||||||||| ||      ||| |||    | || | ||  || ||||||||  |
Db    746 TGAACTTGAACATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGG 805

Qy   1030 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 1089
          | || ||  | || ||||||  |||||| || || |||||  |  |||  || |||||||
Db    806 TTCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAGCTTTTGACT 865

Qy   1090 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1149
          |||| ||       | ||  ||    | |  |||||   |  |||      ||||
Db    866 GGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGCATGCCTCCCTACTACAACC 925

Qy   1150 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1209
          | | ||| |||   || || |  ||| ||||||  || ||  | |||| |||  |  | |
Db    926 TGACAGACATGCATGTGCCAATCGCAGTGTGGAACGGTGGCAACGACTTGCTGGCCGACC 985

Qy   1210 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1269
          |  | || ||  |  | |||| ||  || |  |||| ||||| ||||| | ||| ||||
Db    986 CTCACGATGTTGACCTTTTGCTTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTC 1045

Qy   1270 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1329
          ||   |    |||| |||| || |||||||   ||||||| |||||     | |||||||
Db   1046 CTCCTTACAATCACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATG 1105
```

```
Qy   1330 aaatcatccatctgatg 1346
          ||||  |    |||||
Db   1106 AAATTGTTTCCATGATG 1122
```

RESULT 2
US-09-073-674-2
; Sequence 2, Application US/09073674
; Patent No. 5998189
; GENERAL INFORMATION:
; APPLICANT: Blanchard, Claire
; APPLICANT: Benicourt, Claude
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: Recombinant Dog Gastric Lipase
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Warner-Lambert Company
; STREET: 2800 Plymouth Road
; CITY: Ann Arbor
; STATE: Michigan
; COUNTRY: U.S.A.
; ZIP: 48105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073,674
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Crissey, Todd M.
; REGISTRATION NUMBER: 37,807
; REFERENCE/DOCKET NUMBER: 5072-D1-66-TMC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 734 622-7530
; TELEFAX: 734 622-1553
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1137 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1137
US-09-073-674-2

Query Match 21.7%; Score 444.2; DB 2; Length 1137;
Best Local Similarity 62.8%; Pred. No. 1.1e-117;
Matches 689; Conservative 0; Mismatches 408; Indels 0; Gaps 0;

```
Qy    250 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 309
          |||| ||||    |||||||||| ||| | || |||     |    || || ||   |||||
Db     26 ACCCTGAAGTGACCATGAATATAAGTCAGATGATCACCTACTGGGGATACCCAGCTGAGG 85

Qy    310 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 369
          |||||||||| |  || ||||| || |||||||||     |  |||| ||||||   ||
Db     86 AATATGAAGTTGTGACCGAAGACGGTTATATCCTTGGGATCGACAGAATTCCTTATGGGA 145

Qy    370 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 429
             |  | |||| | |||    || ||||| |  ||  |||| || ||  |  | |
Db    146 GGAAAAATTCAGAGAATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCG 205

Qy    430 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 489
          |   || |  |||||||| |||||||||||||||| ||||||| |||||  | |  |
Db    206 CATCAGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCCTGGCCG 265

Qy    490 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 549
```

```
Db     266 ACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACCTGGGCCAGGAGGAATC 325

Qy     550 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 609
Db     326 TGTACTACTCGCCCGACTCCGTCGAATTCTGGGCTTTCAGCTTTGACGAGATGGCTAAAT 385

Qy     610 ttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 669
Db     386 ATGACCTTCCCGCCACCATTGACTTCATCTTGAAGAAAACGGGACAGGACAAGCTACACT 445

Qy     670 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 729
Db     446 ACGTTGGCCATTCCCAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGC 505

Qy     730 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 789
Db     506 TGGCGAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGTACACCG 565

Qy     790 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 849
Db     566 AAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTCAAGCTTATATTTGGAA 625

Qy     850 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 909
Db     626 ACAAAATATTCTACCCACACCACTTCTTTGATCAATTTCTCGCCACCGAGGTATGCTCCC 685

Qy     910 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 969
Db     686 GCGAGACGGTGGATCTCCTCTGCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTA 745

Qy     970 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 1029
Db     746 TGAACTTGAACATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGG 805

Qy    1030 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 1089
Db     806 TTCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAGCTTTTGACT 865

Qy    1090 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1149
Db     866 GGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGCATGCCTCCCTACTACAACC 925

Qy    1150 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1209
Db     926 TGACAGACATGCATGTGCCAATCGCAGTGTGGAACGGTGGCAACGACTTGCTGGCCGACC 985

Qy    1210 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1269
Db     986 CTCACGATGTTGACCTTTGCTTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTC 1045

Qy    1270 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1329
Db    1046 CTCCTTACAATCACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATG 1105

Qy    1330 aaatcatccatctgatg 1346
Db    1106 AAATTGTTTCCATGATG 1122
```

```
RESULT   3
US-08-227-108-4
; Sequence 4, Application US/08227108
; Patent No. 5807726
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/227,108
;      FILING DATE:  03-APR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fanucci, Allan A.
;      REGISTRATION NUMBER:  30,256
;      REFERENCE/DOCKET NUMBER:  7620-033
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 790-9090
;      TELEFAX:  212 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1140 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1140
US-08-227-108-4
```

```
  Query Match             21.7%;  Score 444.2;  DB 1;  Length 1140;
  Best Local Similarity   62.8%;  Pred. No. 1.1e-117;
  Matches  689;  Conservative    0;  Mismatches  408;  Indels    0;  Gaps    0;

Qy     250 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 309
Db      29 ACCCTGAAGTGACCATGAATATAAGTCAGATGATCACCTACTGGGGATACCCAGCTGAGG 88

Qy     310 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 369
Db      89 AATATGAAGTTGTGACCGAAGACGGTTATATCCTTGGGATCGACAGAATTCCTTATGGGA 148

Qy     370 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 429
Db     149 GGAAAAATTCAGAGAATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCG 208

Qy     430 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 489
Db     209 CATCAGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCCTGGCCG 268

Qy     490 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 549
Db     269 ACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACCTGGGCCAGGAGGAATC 328

Qy     550 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 609
Db     329 TGTACTACTCGCCCGACTCCGTCGAATTCTGGGCTTTCAGCTTTGACGAGATGGCTAAAT 388

Qy     610 ttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 669
Db     389 ATGACCTTCCCGCCACCATTGACTTCATCTTGAAGAAAACGGGACAGGACAAGCTACACT 448

Qy     670 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 729
Db     449 ACGTTGGCCATTCCCAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGC 508

Qy     730 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 789
Db     509 TGGCGAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGTACACCG 568
```

```
Qy     790 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 849
           ||| ||     | |||| |  || |  | ||     |  |  |||||    | ||||| |
Db     569 AAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTCAAGCTTATATTTGGAA 628

Qy     850 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 909
           | |||  |||       ||   |   |||      |   || |  |   | :| ||   ||
Db     629 ACAAAATATTCTACCCACACCACTTCTTTGATCAATTTCTCGCCACCGAGGTATGCTCCC 688

Qy     910 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 969
           | ||  | ||||   | || || ||   | ||||  |  |  |||||||  |||| |
Db     689 GCGAGACGGTGGATCTCCTCTGCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTA 748

Qy     970 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 1029
           ||  |||||||||| ||      ||| |||    | || | ||  || ||||||||  |
Db     749 TGAACTTGAACATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGG 808

Qy    1030 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 1089
           | || ||  | || ||||||  |||||| || || |||||  |  |||  || |||||||
Db     809 TTCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAGCTTTTGACT 868

Qy    1090 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1149
           |||| ||      | ||  ||    | |  |||||   |  |||      ||||
Db     869 GGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGCATGCCTCCCTACTACAACC 928

Qy    1150 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1209
           | | ||| |||   || || |  ||| ||||||  || ||  | |||| ||| |  | | |
Db     929 TGACAGACATGCATGTGCCAATCGCAGTGTGGAACGGTGGCAACGACTTGCTGGCCGACC 988

Qy    1210 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1269
           |  | || ||  |  | ||||| ||  || |  |||| ||||| ||||| | ||| ||||
Db     989 CTCACGATGTTGACCTTTTGCTTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTC 1048

Qy    1270 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1329
           ||   |    |||| |||| || |||||||   |||||| |||||     | ||||||
Db    1049 CTCCTTACAATCACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATG 1108

Qy    1330 aaatcatccatctgatg 1346
           ||||  |     |||||
Db    1109 AAATTGTTTCCATGATG 1125


RESULT   4
US-09-073-674-4
; Sequence 4, Application US/09073674
; Patent No. 5998189
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Warner-Lambert Company
;      STREET:  2800 Plymouth Road
;      CITY:  Ann Arbor
;      STATE:  Michigan
;      COUNTRY:  U.S.A.
;      ZIP:  48105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/073,674
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crissey, Todd M.
;      REGISTRATION NUMBER:  37,807
;      REFERENCE/DOCKET NUMBER:  5072-D1-66-TMC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  734 622-7530
;      TELEFAX:  734 622-1553
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1140 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1140
US-09-073-674-4


  Query Match             21.7%;  Score 444.2;  DB 2;  Length 1140;
  Best Local Similarity   62.8%;  Pred. No. 1.1e-117;
  Matches  689;  Conservative    0;  Mismatches  408;  Indels    0;  Gaps    0;

Qy     250 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 309
           |||  ||||    |||||||||| ||| | || |||    |    || || ||   |||||
Db      29 ACCCTGAAGTGACCATGAATATAAGTCAGATGATCACCTACTGGGGATACCCAGCTGAGG 88

Qy     310 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 369
           |||  |||||| |  || ||||| || |||||||||    |  ||||  ||||||   ||
Db      89 AATATGAAGTTGTGACCGAAGACGGTTATATCCTTGGGATCGACAGAATTCCTTATGGGA 148

Qy     370 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 429
               |  |||| | |||    || ||||| |  ||  |||| || ||  | | |
Db     149 GGAAAAATTCAGAGAATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCG 208

Qy     430 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 489
           |   || |  |||||||| |||||||||||||||| |||||| |||||| ||||| |
Db     209 CATCAGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCCTGGCCG 268

Qy     490 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 549
           | || || |  |||||||| |||||||||||||||| ||| ||||| |  | |   |
Db     269 ACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACCTGGGCCAGGAGGAATC 328

Qy     550 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 609
           |     |||    |  || ||||||||||||| | ||| ||||||||||  |
Db     329 TGTACTACTCGCCCGACTCCGTCGAATTCTGGGCTTTCAGCTTTGACGAGATGGCTAAAT 388

Qy     610 ttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaaagatctatt 669
           |||||||| ||    ||  |||| || ||| |||||||||| ||||| ||| |  | |
Db     389 ATGACCTTCCCGCCACCATTGACTTCATCTTGAAGAAAACGGGACAGGACAAGCTACACT 448

Qy     670 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 729
           | || ||| |||| ||||||||||||||| || || || || ||||||||||  ||  |||
Db     449 ACGTTGGCCATTCCCAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGC 508

Qy     730 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 789
           ||||  |    |||||||  |  | ||| |||| ||| | ||||| || ||| |  |
Db     509 TGGCGAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGTACACCG 568

Qy     790 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 849
           ||| ||     | |||| |  || |  | ||     |  |  |||||    | ||||| |
Db     569 AAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTCAAGCTTATATTTGGAA 628

Qy     850 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 909
           | |||  |||       ||   |   |||      |   || |  |   | :| ||   ||
Db     629 ACAAAATATTCTACCCACACCACTTCTTTGATCAATTTCTCGCCACCGAGGTATGCTCCC 688

Qy     910 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 969
           | ||  | ||||   | || || ||   | ||||  |  |  |||||||  |||| |
Db     689 GCGAGACGGTGGATCTCCTCTGCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTA 748

Qy     970 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 1029
           ||  |||||||||| ||      ||| |||    | || | ||  || ||||||||  |
Db     749 TGAACTTGAACATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGG 808

Qy    1030 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 1089
```

```
Db    809 TTCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAGCTTTTGACT 868

Qy   1090 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1149
Db    869 GGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGCATGCCTCCCTACTACAACC 928

Qy   1150 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1209
Db    929 TGACAGACATGCATGTGCCAATCGCAGTGTGGAACGGTGGCAACGACTTGCTGGCCGACC 988

Qy   1210 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1269
Db    989 CTCACGATGTTGACCTTTTGCTTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTC 1048

Qy   1270 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1329
Db   1049 CTCCTTACAATCACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATG 1108

Qy   1330 aaatcatccatctgatg 1346
Db   1109 AAATTGTTTCCATGATG 1125


RESULT   5
US-08-227-108-6
; Sequence 6, Application US/08227108
; Patent No. 5807726
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/227,108
;      FILING DATE:  03-APR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fanucci, Allan A.
;      REGISTRATION NUMBER:  30,256
;      REFERENCE/DOCKET NUMBER:  7620-033
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 790-9090
;      TELEFAX:  212 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1146 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
US-08-227-108-6


  Query Match           21.7%;  Score 444.2;  DB 1;  Length 1146;
  Best Local Similarity  62.8%;  Pred. No. 1.1e-117;
  Matches  689;  Conservative    0;  Mismatches  408;  Indels    0;  Gaps    0;

Qy    250 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 309
Db     26 ACCCTGAAGTGACCATGAATATAAGTCAGATGATCACCTACTGGGGATACCCAGCTGAGG 85

Qy    310 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 369
Db     86 AATATGAAGTTGTGACCGAAGACGGTTATATCCTTGGGATCGACAGAATTCCTTATGGGA 145

Qy    370 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 429
Db    146 GGAAAAATTCAGAGAATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCG 205

Qy    430 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 489
Db    206 CATCAGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCCTGGCCG 265

Qy    490 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 549
Db    266 ACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACCTGGGCCAGGAGGAATC 325

Qy    550 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 609
Db    326 TGTACTACTCGCCCGACTCCGTCGAATTCTGGGCTTTCAGCTTTGACGAGATGGCTAAAT 385

Qy    610 ttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 669
Db    386 ATGACCTTCCCGCCACCATTGACTTCATCTTGAAGAAAACGGGACAGGACAAGCTACACT 445

Qy    670 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 729
Db    446 ACGTTGGCCATTCCCAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGC 505

Qy    730 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 789
Db    506 TGGCGAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGTACACCG 565

Qy    790 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 849
Db    566 AAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTCAAGCTTATATTTGGAA 625

Qy    850 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccttttgtggcc 909
Db    626 ACAAAATATTCTACCCACACCACTTCTTTGATCAATTTCTCGCCACCGAGGTATGCTCCC 685

Qy    910 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 969
Db    686 GCGAGACGGTGGATCTCCTCTGCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTA 745

Qy    970 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 1029
Db    746 TGAACTTGAACATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGG 805

Qy   1030 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 1089
Db    806 TTCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAGCTTTTGACT 865

Qy   1090 ggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1149
Db    866 GGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGCATGCCTCCCTACTACAACC 925

Qy   1150 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1209
Db    926 TGACAGACATGCATGTGCCAATCGCAGTGTGGAACGGTGGCAACGACTTGCTGGCCGACC 985

Qy   1210 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1269
Db    986 CTCACGATGTTGACCTTTTGCTTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTC 1045

Qy   1270 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1329
Db   1046 CTCCTTACAATCACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATG 1105

Qy   1330 aaatcatccatctgatg 1346
```

```
Db   1106 AAATTGTTTCCATGATG 1122


RESULT   6
US-09-073-674-6
; Sequence 6, Application US/09073674
; Patent No. 5998189
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Warner-Lambert Company
;      STREET:  2800 Plymouth Road
;      CITY:  Ann Arbor
;      STATE:  Michigan
;      COUNTRY:  U.S.A.
;      ZIP:  48105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/073,674
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crissey, Todd M.
;      REGISTRATION NUMBER:  37,807
;      REFERENCE/DOCKET NUMBER:  5072-D1-66-TMC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  734 622-7530
;      TELEFAX:  734 622-1553
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1146 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
US-09-073-674-6


  Query Match           21.7%;  Score 444.2;  DB 2;  Length 1146;
  Best Local Similarity  62.8%;  Pred. No. 1.1e-117;
  Matches  689;  Conservative    0;  Mismatches  408;  Indels    0;  Gaps    0;

Qy    250 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 309
          |||| ||||    |||||||| ||| | || |||     |    || || ||   |||||
Db     26 ACCCTGAAGTGACCATGAATATAAGTCAGATGATCACCTACTGGGGATACCCAGCTGAGG 85

Qy    310 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 369
          |||||||||| |  || ||||| || |||||||||     |  |||| ||||||   ||
Db     86 AATATGAAGTTGTGACCGAAGACGGTTATATCCTTGGGATCGACAGAATTCCTTATGGGA 145

Qy    370 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 429
             |  | |||| | |||   || ||||| |  ||  |||| || ||  |  | |
Db    146 GGAAAAATTCAGAGAATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCG 205

Qy    430 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 489
          |   || |  |||||||| |||||||||||||||||| ||||||| |||||| |||||
Db    206 CATCAGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCCTGGCCG 265

Qy    490 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 549
          | || || |  |||||||| |||||||||||||||| ||| ||||| |  | |   |
Db    266 ACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACCTGGGCCAGGAGGAATC 325

Qy    550 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 609
             |     |||    |||    |  || |||||||||||||| | ||| ||||||||||  |

Db    326 TGTACTACTCGCCCGACTCCGTCGAATTCTGGGCTTTCAGCTTTGACGAGATGGCTAAAT 385

Qy    610 ttgaccttcctgcagtgataaactttatttttgcagaaaacgggccaggaaaagatctatt 669
          |||||||||| ||    ||  |||| || ||| ||||||||||| ||||| ||| |  | |
Db    386 ATGACCTTCCCGCCACCATTGACTTCATCTTGAAGAAAACGGGACAGGACAAGCTACACT 445

Qy    670 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 729
          | || ||| |||| |||||||||||||| || || || || |||||||||  ||  |||
Db    446 ACGTTGGCCATTCCCAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGC 505

Qy    730 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 789
          ||||  |    ||||||| |  | ||| ||||| ||| | ||||| || ||| |  |
Db    506 TGGCGAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGTACACCG 565

Qy    790 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 849
          ||| ||     | |||| |  || |  | ||    |  |  |||||    | |||||| |
Db    566 AAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTCAAGCTTATATTTGGAA 625

Qy    850 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccttttgtggcc 909
          | |||  |||        ||   |  |||      |   ||  |  |  | | ||  ||
Db    626 ACAAAATATTCTACCCACACCACTTCTTTGATCAATTTCTCGCCACCGAGGTATGCTCCC 685

Qy    910 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 969
          | ||  | ||||   | || || ||   | ||||  |  |  |||||||  |||| |
Db    686 GCGAGACGGTGGATCTCCTCTGCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTA 745

Qy    970 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 1029
            ||  |||||||||| ||       ||| |||     | || | ||  || |||||||| |
Db    746 TGAACTTGAACATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGG 805

Qy   1030 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 1089
          | || ||  | || ||||||  ||||||| || || |||||  |  |||  || |||||||
Db    806 TTCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAGCTTTTGACT 865

Qy   1090 ggggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1149
          ||||  ||        | ||  ||     | |  |||||  |  |||          ||||
Db    866 GGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGCATGCCTCCCTACTACAACC 925

Qy   1150 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1209
          | |  || |||    || || |  ||| |||||||  || ||  | |||| |||  |  | |
Db    926 TGACAGACATGCATGTGCCAATCGCAGTGTGGAACGGTGGCAACGACTTGCTGGCCGACC 985

Qy   1210 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1269
          |  | || ||  |  | |||| ||  || |  |||| ||||| ||||| | ||| ||||
Db    986 CTCAAGATGTTGACCTTTTGCTTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTC 1045

Qy   1270 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1329
          ||   |    |||| |||| || |||||||  |||||| |||||      | |||||||
Db    1046 CTCCTTACAATCACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATG 1105

Qy   1330 aaatcatccatctgatg 1346
          ||||  |     |||||
Db    1106 AAATTGTTTCCATGATG 1122


RESULT   7
US-08-227-108-1
; Sequence 1, Application US/08227108
; Patent No. 5807726
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
```

```
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/227,108
;      FILING DATE:  03-APR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fanucci, Allan A.
;      REGISTRATION NUMBER:  30,256
;      REFERENCE/DOCKET NUMBER:  7620-033
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 790-9090
;      TELEFAX:  212 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1528 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
US-08-227-108-1


  Query Match             21.7%;  Score 444.2;  DB 1;  Length 1528;
  Best Local Similarity   62.8%;  Pred. No. 1.3e-117;
  Matches  689;  Conservative    0;  Mismatches  408;  Indels    0;  Gaps    0;

Qy    250 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 309
          |||| ||||    |||||||| ||| | || |||    |    || || ||   |||||
Db     26 ACCCTGAAGTGACCATGAATATAAGTCAGATGATCACCTACTGGGGATACCCAGCTGAGG 85

Qy    310 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 369
          |||||||||| |  || ||||| || |||||||||    |  |||| ||||||   ||
Db     86 AATATGAAGTTGTGACCGAAGACGGTTATATCCTTGGGATCGACAGAATTCCTTATGGGA 145

Qy    370 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 429
             |  | |||| | |||    || ||||| |  ||  |||| || ||  |  | |
Db    146 GGAAAAATTCAGAGAATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCG 205

Qy    430 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 489
          |   || |  |||||||| |||||||||||||||||| ||||||| |||||| ||||| |
Db    206 CATCAGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCCTGGCCG 265

Qy    490 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 549
          | || || |  |||||||| ||||||||||||||| ||| ||||| | | |   |
Db    266 ACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACCTGGGCCAGGAGGAATC 325

Qy    550 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 609
          |     |||    |||   |  || |||||||||||||| | ||| ||||||||||  |
Db    326 TGTACTACTCGCCCGACTCCGTCGAATTCTGGGCTTTCAGCTTTGACGAGATGGCTAAAT 385

Qy    610 ttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 669
           |||||||||  ||    ||  |||| || ||| |||||||||| ||||| ||| |  | |
Db    386 ATGACCTTCCCGCCACCATTGACTTCATCTTGAAGAAAACGGGACAGGACAAGCTACACT 445

Qy    670 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 729
          | || |||  |||| |||||||||||||| || || || || ||||||||||   ||  |||
Db    446 ACGTTGGCCATTCCCAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGC 505

Qy    730 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 789
          ||||  |    |||||||  |  | ||| ||||| ||| | ||||| || ||| |  |
Db    506 TGGCGAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGTACACCG 565

Qy    790 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 849
          ||| ||     | |||| |  || |  | ||     |  |  |||||    | ||||| |
Db    566 AAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTCAAGCTTATATTTGGAA 625

Qy    850 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccttttgtggcc 909
          | |||  |||        ||   |  |||     |   || |  |   |  | ||  ||
Db    626 ACAAAATATTCTACCCACACCACTTCTTTGATCAATTTCTCGCCACCGAGGTATGCTCCC 685

Qy    910 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 969
          | ||   | ||||   | || || ||   | ||||  |  |  |||||||  |||| |
Db    686 GCGAGACGGTGGATCTCCTCTGCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTA 745

Qy    970 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 1029
          ||  |||||||||| ||      ||| |||    | || | ||  || |||||||| |
Db    746 TGAACTTGAACATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGG 805

Qy   1030 tgcaaaatattctacactggagccaggcagtgaatttctggtgaactccgggcatttgact 1089
          | || ||  | || ||||||  ||||||| || || ||||| .|  |||  || |||||||
Db    806 TTCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAGCTTTTGACT 865

Qy   1090 gggcgagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1149
          |||| ||         | ||  ||    | |   |||||    |  |||      ||||
Db    866 GGGCAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGCATGCCTCCCTACTACAACC 925

Qy   1150 tcacagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1209
          | | ||| |||   || || |  ||| ||||||  || ||  | ||||  |||  |  | |
Db    926 TGACAGACATGCATGTGCCAATCGCAGTGTGGAACGGTGGCAACGACTTGCTGGCCGACC 985

Qy   1210 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1269
          |  | || ||  |  |  |||| ||  || |  |||| ||||| ||||| | ||| ||||
Db    986 CTCACGATGTTGACCTTTTGCTTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTC 1045

Qy   1270 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1329
          ||   |    |||| |||| || |||||||   |||||| |||||     | |||||||
Db    1046 CTCCTTACAATCACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATG 1105

Qy   1330 aaatcatccatctgatg 1346
          .. |||| |    |||||
Db    1106 AAATTGTTTCCATGATG 1122


RESULT   8
US-09-073-674-1
; Sequence 1, Application US/09073674
; Patent No. 5998189
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Warner-Lambert Company
;      STREET:  2800 Plymouth Road
;      CITY:  Ann Arbor
;      STATE:  Michigan
;      COUNTRY:  U.S.A.
;      ZIP:  48105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/073,674
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crissey, Todd M.
;      REGISTRATION NUMBER:  37,807
;      REFERENCE/DOCKET NUMBER:  5072-D1-66-TMC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  734 622-7530
;      TELEFAX:  734 622-1553
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1528 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
```

```
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
US-09-073-674-1


  Query Match             21.7%;  Score 444.2;  DB 2;  Length 1528;
  Best Local Similarity   62.8%;  Pred. No. 1.3e-117;
  Matches  689;  Conservative    0;  Mismatches  408;  Indels    0;  Gaps    0;

Qy    250 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 309
          |||| ||||    |||||||||| ||| | || |||    |    || || ||   |||||
Db     26 ACCCTGAAGTGACCATGAATATAAGTCAGATGATCACCTACTGGGGATACCCAGCTGAGG 85

Qy    310 aatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctcgaggcc 369
          |||||||||| |  || ||||| || |||||||||    |  |||| ||||||    ||
Db     86 AATATGAAGTTGTGACCGAAGACGGTTATATCCTTGGGATCGACAGAATTCCTTATGGGA 145

Qy    370 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 429
          |  | |||| | |||    || ||||| |  ||  |||| || ||  |  | |
Db    146 GGAAAAATTCAGAGAATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCG 205

Qy    430 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 489
          |   ||  | |||||||| |||||||||||||||||| ||||||| |||||| ||||| |
Db    206 CATCAGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCCTGGCCG 265

Qy    490 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 549
          | || || |  |||||||| ||||||||||||||| ||| ||||| |  | |   |
Db    266 ACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACCTGGGCCAGGAGGAATC 325

Qy    550 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 609
          |    |||    |||   |  || ||||||||||||| | ||| ||||||||||  |
Db    326 TGTACTACTCGCCCGACTCCGTCGAATTCTGGGCTTTCAGCTTTGACGAGATGGCTAAAT 385

Qy    610 ttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 669
           |||||||||| ||    ||  |||| || ||| |||||||||| ||||| ||| |  | |
Db    386 ATGACCTTCCCGCCACCATTGACTTCATCTTGAAGAAAACGGGACAGGACAAGCTACACT 445

Qy    670 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 729
          | || ||| |||| ||||||||||||| || || || || ||||||||||  ||  |||
Db    446 ACGTTGGCCATTCCCAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGC 505

Qy    730 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 789
          ||||  |    |||||||  |  | ||| ||||| ||| | ||||| || ||| |  |
Db    506 TGGCGAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGTACACCG 565

Qy    790 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 849
          ||| ||     | |||| |  || |  | ||     |  |  |||||    | ||||| |
Db    566 AAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTCAAGCTTATATTTGGAA 625

Qy    850 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccttgtggcc 909
          | ||| .|||       ||  |   |||      |   || |  |    |  ||   ||
Db    626 ACAAAATATTCTACCCACACCACTTCTTTGATCAATTTCTCGCCACCGAGGTATGCTCCC 685

Qy    910 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 969
          |  ||   | ||||   | || || ||   | ||||  |  |  |||||||  |||| |
Db    686 GCGAGACGGTGGATCTCCTCTGCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTA 745

Qy    970 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 1029
            ||  |||||||||| ||      ||| |||    | || | ||  || |||||||| |
Db    746 TGAACTTGAACATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGG 805

Qy   1030 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 1089
          | || ||  | || |||||| |||||| || || ||||  |  |||  || ||||||||
Db    806 TTCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAGCTTTTGACT 865

Qy   1090 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1149
          |||| ||         | ||  ||    | |||||  |  |||      ||||
Db    866 GGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGCATGCCTCCCTACTACAACC 925

Qy   1150 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1209
          | | ||| |||   || || | ||| ||||||  || ||  | |||| |||  |  | |
Db    926 TGACAGACATGCATGTGCCAATCGCAGTGTGGAACGGTGGCAACGACTTGCTGGCCGACC 985

Qy   1210 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1269
          |  | || ||  |  |  |||| ||  || |  |||| ||||| ||||| | ||| ||||
Db    986 CTCACGATGTTGACCTTTTGCTTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTC 1045

Qy   1270 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1329
          ||   |    |||| |||| || |||||||   |||||| |||||     | |||||||
Db   1046 CTCCTTACAATCACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATG 1105

Qy   1330 aaatcatccatctgatg 1346
          ||||  |     |||||
Db   1106 AAATTGTTTCCATGATG 1122


RESULT   9
US-08-751-782-2
; Sequence 2, Application US/08751782
; Patent No. 5821352
;  GENERAL INFORMATION:
;    APPLICANT:  Heintz, Nathaniel
;    APPLICANT:  Gubbay, Johnathan
;    APPLICANT:  Skinner, Michael
;    TITLE OF INVENTION:  A cDNA Library Prepared during
;    TITLE OF INVENTION:  Regression of Rat Prostate and Methods of Use Thereof
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  David A. Jackson, Esq.
;      STREET:  411 Hackensack Ave, Continental Plaza, 4th
;      STREET:  Floor
;      CITY:  Hackensack
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/751,782
;      FILING DATE:  18-NOV-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jackson Esq., David A.
;      REGISTRATION NUMBER:  26,742
;      REFERENCE/DOCKET NUMBER:  600-1-190
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-487-5800
;      TELEFAX:  201-343-1684
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  178 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    IMMEDIATE SOURCE:
;      CLONE:  10.2
US-08-751-782-2


  Query Match              2.4%;  Score 48.2;  DB 1;  Length 178;
  Best Local Similarity   55.8%;  Pred. No. 0.00013;
  Matches   92;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy    541 gaaaacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgaga 600
          | |||||   |   ||   |  |||      | || || |||| ||| |||| | ||||
Db      8 GGAAACATGTGCGCCTAGACCCAGATTCTAAAGAATTTTGGGATTTTAGTTTTAATGAAC 67

Qy    601 tggctaggtttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaa 660
                |  ||||| || ||  | ||  | || ||| || |  ||||  | ||   |
Db     68 AAATAGAATACGACCTCCCAGCCATCATTTATTTCATTCTGAATGAAACAAGACAAACAC 127
```

```
Qy    661 agatctattatgtcggctattcacagggcaccaccatgggcttta 705
          | || || ||| | ||| |||| ||||||  |    || |  | |
Db    128 AAATTTACTATATTGGCCATTCCCAGGGCGTCTATCTGCGTATCA 172


RESULT 10
US-08-925-171-2
; Sequence 2, Application US/08925171
; Patent No. 5928871
;  GENERAL INFORMATION:
;    APPLICANT:  Heintz, Nathaniel
;    APPLICANT:  Gubbay, Johnathan
;    APPLICANT:  Skinner, Michael
;    TITLE OF INVENTION:  A cDNA Library Prepared during
;    TITLE OF INVENTION:  Regression of Rat Prostate and Methods of Use Thereof
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  David A. Jackson, Esq.
;      STREET:  411 Hackensack Ave, Continental Plaza, 4th
;      STREET:  Floor
;      CITY:  Hackensack
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/925,171
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jackson Esq., David A.
;      REGISTRATION NUMBER:  26,742
;      REFERENCE/DOCKET NUMBER:  600-1-190 CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-487-5800
;      TELEFAX:  201-343-1684
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  178 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    IMMEDIATE SOURCE:
;      CLONE:  10.2
US-08-925-171-2


  Query Match             2.4%;  Score 48.2;  DB 2;  Length 178;
  Best Local Similarity   55.8%;  Pred. No. 0.00013;
  Matches   92;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy    541 gaaaacacaagacactctccatagaccaagatgagttctgggctttcagttatgatgaga 600
          | |||||   |   ||   |  |||     | || || |||| ||| |||| | ||||
Db      8 GGAAACATGTGCGCCTAGACCCAGATTCTAAAGAATTTTGGGATTTTAGTTTTAATGAAC 67

Qy    601 tggctaggtttgaccttcctgcagtgataaactttatttttgcagaaaacgggccaggaaa 660
            |  ||||| || ||  | ||  | || ||| || |  |||| | ||   |
Db     68 AAATAGAATACGACCTCCCAGCCATCATTTATTTCATTCTGAATGAAACAAGACAAACAC 127

Qy    661 agatctattatgtcggctattcacagggcaccaccatgggcttta 705
          | || || ||| | ||| |||| ||||||  |    || |  | |
Db    128 AAATTTACTATATTGGCCATTCCCAGGGCGTCTATCTGCGTATCA 172


RESULT 11
US-09-356-952-12
; Sequence 12, Application US/09356952
; Patent No. 6117663
; GENERAL INFORMATION:
;  APPLICANT: Boriack-Sjodin, Ann
;  APPLICANT: Margarit, S. M.
;  APPLICANT: Bor-Sogi, Dafna
;  APPLICANT: Cole, Philip
;  APPLICANT: Kuriyan, John
;  TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE
;  TITLE OF INVENTION: THEREOF
;  FILE REFERENCE: 600-1-228N
;  CURRENT APPLICATION NUMBER: US/09/356,952
;  CURRENT FILING DATE: 1999-07-19
;  EARLIER APPLICATION NUMBER: 60/093,631
;  EARLIER FILING DATE: 1998-07-21
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
;   LENGTH: 43676
;   TYPE: DNA
;   ORGANISM: Saccharomyces cerevisiae
US-09-356-952-12


  Query Match             2.3%;  Score 46.4;  DB 3;  Length 43676;
  Best Local Similarity   46.0%;  Pred. No. 0.0079;
  Matches  232;  Conservative    0;  Mismatches  266;  Indels    6;  Gaps    2;

Qy    219 aattcagtacatatgccaactaaagctgtggacccagaagcattcatgaatattagtgaa 278
          ||| || |||| |  | ||    | | |  |||     |  | |   | || ||| | |
Db  14682 aatccattacagaatctaatcttatcggaagacttgacattagttgcggatcttaattat 14741

Qy    279 atcatccaacatcaaggctatccctgtgaggaatatgaagtcgcaactgaagatgggt-- 336
              |  | ||  |      ||     |||||| |  | | |  || |||||||||
Db  14742 tattttaaccagtacaatattcaaatagaggaatttagattggagaccgaagatgggttt 14801

Qy    337 -atatcctttctgttaacaggattcctcgaggcctagtgcaacctaagaagacaggttcc 395
           |||   ||      ||  ||| ||   |     |  |  |  |   | |
Db  14802 gttatagatttgtggcacttgataccaaaatatagaacgacagattctgacaagaagaag 14861

Qy    396 aggcctgtggtgttactgcagcatggcctagttggaggtgctagcaactggatttccaac 455
          |||||     | ||  |||  |||||||| |   | |   | |    |     |||||
Db  14862 aggccacccattttgatgctacatggccttttgcaaagcagtggttcgttcgcatccaa- 14920

Qy    456 ctgcccaacaatagcctgggcttcattctggcagatgctggttttgacgtgtggatgggg 515
              |  ||    |||| |   |||||    |  | |||| ||| | ||| | |||
Db  14921 --tcgtagaaaatctctggcatattttctgtatcaatccggttacgacatatggttaggg 14978

Qy    516 aacagcaggggaaacgcctggtctcgaaaacacaagacactctccatagaccaagatgag 575
          || | |||  |        || |   |    || |  |  ·          | || |
Db  14979 aataacagatgcgggtttaggccggaatggaacgaagcgaaagtaccgacactagcttcc 15038

Qy    576 ttctgggctttcagttatgatgagatggctaggtttgaccttcctgcagtgataaacttt 635
            ||||  |       |   ||||||| || ||  || ||  |     ||||  |
Db  15039 aggtgggactgggaccttcgcgagatggttaagtacgatctgacccttttgattgatacc 15098

Qy    636 attttgcagaaaacgggccaggaaaagatctattatgtcggctattcacagggcaccacc 695
            | ||    || |||       |||||| |   ||   ||    |||| |||||||| ||
Db  15099 gtgttagctaagacgcagtttgaaaagcttactttgatctcgcattctcagggcactaca 15158

Qy    696 atgcgctttattgcattttccacc 719
            ||| ||||| |  ||   || |
Db  15159 cagcggtttatgggcttggtcaac 15182


RESULT 12
US-08-765-081-1/c
; Sequence 1, Application US/08765081
; Patent No. 5798260
;  GENERAL INFORMATION:
;    APPLICANT:  Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
;    TITLE OF INVENTION:  Escherichia Coli 0157:H7 Epithelial Adhesin
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
```

```
;       ADDRESSEE:  Christensen, O'Connor, Johnson and Kindness
;       STREET:  2800 Pacific First Center, 1420 Fifth Avenue
;       CITY:  Seattle
;       STATE:  Washington
;       COUNTRY:  USA
;       ZIP:  98101-2347
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette-3.5 inch, 1.44Mb storage
;       COMPUTER:  IBM PC compatible/Pentium
;       OPERATING SYSTEM:  MS-Window 3.1
;       SOFTWARE:  Word for Windows-6.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/765,081
;       FILING DATE:  March 26, 1997
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/06994
;       FILING DATE:  June 7, 1995
;       APPLICATION NUMBER:  US 08/265,714
;       FILING DATE:  June 24, 1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Shelton, Dennis K.
;       REGISTRATION NUMBER:  26,997
;       REFERENCE/DOCKET NUMBER:  CHOR-1-10286
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  1-206-682-8100; 1-206-224-0718 (direct)
;       TELEFAX:  1-206-224-0779
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  8041 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Escherichia coli O157:H7
;       STRAIN:  86-24 NALR
;     IMMEDIATE SOURCE:
;       CLONE:  pEAR
US-08-765-081-1


  Query Match             2.1%;  Score 42;  DB 1;  Length 8041;
  Best Local Similarity   53.4%;  Pred. No. 0.058;
  Matches   87;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy      33 aaaatatgtgaagagtttttaaacccacaaattcttcttactttagaattagttgttaca 92
           ||| ||||||||    |||| || | ||||  |  | | | |||| || ||  ||||| |
Db    1541 AAAGTATGTGAACTTATTTTTAAACAACAAGATAATTTGATTTTAAAAATAAATGTTAGA 1482

Qy      93 ttggcaggaaaaaataaatgcagatgttggaccatgttggaaaccttgtcaagacagtgg 152
           ||  |  ||| || || |   | ||   | ||  |    |  |   |   | | |  |||
Db    1481 TTTCCCAGAATAATTATAAAGAAATACCGAACATTACAAGGTAAAGTCGTATGGCTATGG 1422

Qy     153 attgtctcacacagaatggaaatgtggcttctgattctggtgg 195
           ||  |  ||      || |||  ||   || ||  |    |||
Db    1421 ATATTGNCATTTGTTATTGAATGGTTAATTGTGTGTACTTTGG 1379


RESULT  13
US-09-098-082-1/c
; Sequence 1, Application US/09098082
; Patent No. 6040421
;  GENERAL INFORMATION:
;    APPLICANT:  Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
;    TITLE OF INVENTION:  Escherichia Coli 0157:H7 Epithelial Adhesin
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen, O'Connor, Johnson and Kindness
;      STREET:  2800 Pacific First Center, 1420 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette-3.5 inch, 1.44Mb storage
;      COMPUTER:  IBM PC compatible/Pentium II
;      OPERATING SYSTEM:  MS-Windows 95
;      SOFTWARE:  Word for Windows-6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/098,082
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/765,081
;      FILING DATE:  March 26, 1997
;      APPLICATION NUMBER:  PCT/US95/06994
;      FILING DATE:  June 7, 1995
;      APPLICATION NUMBER:  US 08/265,714
;      FILING DATE:  June 24, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sheiness, Diana K.
;      REGISTRATION NUMBER:  35,356
;      REFERENCE/DOCKET NUMBER:  CHOR-1-12402
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  1-206-682-8100; 1-206-224-0735 (direct)
;      TELEFAX:  1-206-224-0779
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8041 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli O157:H7
;      STRAIN:  86-24 NALR
;    IMMEDIATE SOURCE:
;      CLONE:  pEAR
US-09-098-082-1


  Query Match             2.1%;  Score 42;  DB 3;  Length 8041;
  Best Local Similarity   53.4%;  Pred. No. 0.058;
  Matches   87;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy      33 aaaatatgtgaagagtttttaaacccacaaattcttcttactttagaattagttgttaca 92
           ||| ||||||||    |||| || | ||||  |  | | | |||| || ||  ||||| |
Db    1541 AAAGTATGTGAACTTATTTTTAAACAACAAGATAATTTGATTTTAAAAATAAATGTTAGA 1482

Qy      93 ttggcaggaaaaaataaatgcagatgttggaccatgttggaaaccttgtcaagacagtgg 152
           ||  |  ||| || || |   | ||   | ||  |    |  |   |   | | |  |||
Db    1481 TTTCCCAGAATAATTATAAAGAAATACCGAACATTACAAGGTAAAGTCGTATGGCTATGG 1422

Qy     153 attgtctcacacagaatggaaatgtggcttctgattctggtgg 195
           ||  |  ||      || |||  ||   || ||  |    |||
Db    1421 ATATTGNCATTTGTTATTGAATGGTTAATTGTGTGTACTTTGG 1379


RESULT  14
PCT-US95-06994-1/c
; Sequence 1, Application PC/TUS9506994
;  GENERAL INFORMATION:
;    APPLICANT:  Children's Hospital & Medical Center
;    APPLICANT:  University of Washington
;    APPLICANT:  Washington State University Research Foundation
;    APPLICANT:  TARR, PHILLIP I
;    APPLICANT:  BILGE, SIMA S
;    APPLICANT:  BESSER, THOMAS E
;    APPLICANT:  VARY JR, JAMES C
;    TITLE OF INVENTION:  ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESSPLLC
```

```
;      STREET:  SUITE 2800, 1420 FIFTH AVENUE
;      CITY:  SEATTLE
;      STATE:  WASHINGTON
;      COUNTRY:  USA
;      ZIP:  WA 98101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/06994
;      FILING DATE:  07-JUN-95
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/265,714
;      FILING DATE:  24-JUN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BRODERICK, THOMAS F
;      REGISTRATION NUMBER:  31,332
;      REFERENCE/DOCKET NUMBER:  CHOR-18591
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 682 8100
;      TELEFAX:  (206) 224 0779
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8041 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli O157:H7
;      STRAIN:  86-24 NALR
;    IMMEDIATE SOURCE:
;      CLONE:  pEAR
PCT-US95-06994-1


  Query Match             2.1%;  Score 42;  DB 5;  Length 8041;
  Best Local Similarity   53.4%;  Pred. No. 0.058;
  Matches   87;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy      33 aaaatatgtgaagagtttttaaacccacaaattcttcttactttagaattagttgttaca 92
           ||| ||||||||    |||| || | ||||  |  | | | |||| || ||  ||||| |
Db    1541 AAAGTATGTGAACTTATTTTTAAACAACAAGATAATTTGATTTTAAAAATAAATGTTAGA 1482

Qy      93 ttggcaggaaaaaataaatgcagatgttggaccatgttggaaaccttgtcaagacagtgg 152
           ||  |  ||| || || |   | ||   | ||  |    |  |   |   | | |  |||
Db    1481 TTTCCCAGAATAATTATAAAGAAATACCGAACATTACAAGGTAAAGTCGTATGGCTATGG 1422

Qy     153 attgtctcacacagaatggaaatgtggcttctgattctggtgg 195
           ||  |  ||      || |||  ||   || ||  |    |||
Db    1421 ATATTGNCATTTGTTATTGAATGGTTAATTGTGTGTACTTTGG 1379


RESULT  15
US-08-698-805-5
; Sequence 5, Application US/08698805
; Patent No. 5869288
;  GENERAL INFORMATION:
;    APPLICANT:  Chapman, Martin
;    APPLICANT:  Arruda, L. Karla
;    TITLE OF INVENTION:  Molecular Cloning of Cockroach
;    TITLE OF INVENTION:  Allergens, Amino Acid and Nucleotide Sequences Therefore,
;    TITLE OF INVENTION:  and Recombinant Expression Thereof
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Oblon, Spivak, McClelland, Maier & Neustadt,
;      ADDRESSEE:  P.C.
;      STREET:  1755 S. Jefferson Davis Highway, Suite 400
;      CITY:  Arlington
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/698,805
;      FILING DATE:  16-AUG-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/002,510
;      FILING DATE:  18-AUG-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kelber, Steven B.
;      REGISTRATION NUMBER:  30,073
;      REFERENCE/DOCKET NUMBER:  494-203-27
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-413-3000
;      TELEFAX:  703-413-2220
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1140 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3..602
US-08-698-805-5


  Query Match             2.0%;  Score 40.4;  DB 2;  Length 1140;
  Best Local Similarity   45.5%;  Pred. No. 0.059;
  Matches  143;  Conservative    0;  Mismatches  171;  Indels    0;  Gaps    0;

Qy    1525 gacacttacatttacatttttttttctgtaaattaaagtacttattaggtaaatagaggtt 1584
           ||| || |   |||||||| ||  |   |||| ||| | || |||   | |   |
Db     827 GACGCTGATTCTTACATTTAATTCTTCACAATTTAAGAATTTTTTAACAATAGTAATTAC 886

Qy    1585 ttgcatgctattatatatttctaccatcttgaagggtaggttttacctgatagccagaaaa 1644
             |   | | | || |  |   || |    |   || || |   | ||    | | |
Db     887 GATCAACATTTCAGATCTGTTTAGATATGATTGCAAAGCTTGTTTATAATCAGAAAATGA 946

Qy    1645 tatctagacattctctatatcattcaggtaaatctctttaaaacacctattgtttttttct 1704
             ||   | ||   |   ||    |   || | |   |   | |  |   |||||||
Db     947 CTTCCTAAAATCAACAGCATATGGCGCAAAATTTTTCGTTCTAAATTTCCAGTTTTTTAA 1006

Qy    1705 ataagccatatttttggagcactaaagtaaaatggcaaattgggacagatattgaggtct 1764
           ||     || |||||||   |||     |  | | | || |||    ||||  | ||    |
Db    1007 ATGTATAATTTTTTTGGTAAACTTTTATTTACTAGAAATTTGATCCAGAAGTAGACTGAT 1066

Qy    1765 ggagtctgtggattattgttgactttgacaaaataagctagacattttcaccttgttgcc 1824
             ||  |   ||| | ||   | | | | ||      | |||   | |  |||
Db    1067 AATTTCCTTTACTTACTTTTTGGTATTAAACAAAGTTGGAAACAAAATAATTTTGAAAAA 1126

Qy    1825 acagagacataaca 1838
           | | | | | || |
Db    1127 AAAAAAAAAAAAAA 1140


Search completed: August 24, 2001, 22:38:56
Job time: 7252 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: August 24, 2001, 19:50:13 ; Search time 2518.61 Seconds
(without alignments)
7671.546 Million cell updates/sec

Title: US-09-333-159-45
Perfect score: 2044
Sequence: 1 gtcgacccacgcgtccgggg.....aaaaaaaaaagggcggccgc 2044

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_esthum29:*
177: em_esthum30:*
178: em_esthum31:*
179: em_esthum32:*
180: em_esthum33:*
181: em_estom3:*
182: em_estpl11:*
183: em_estro21:*
184: em_estro22:*
185: em_estro23:*
186: em_htc:*
187: gb_est107:*
188: gb_est108:*
189: gb_est109:*
190: gb_est110:*
191: gb_est111:*
192: gb_htc:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_other:*
207: em_gss_pln1:*
208: em_gss_pln2:*
209: em_gss_pro:*
210: em_gss_rod1:*
211: em_gss_rod2:*
212: em_gss_rod3:*
213: em_gss_rod4:*
214: em_gss_rod5:*
215: em_gss_vrt1:*
216: em_gss_vrt2:*
217: em_gss_vrt3:*
218: gb_gss1:*
219: gb_gss2:*
220: gb_gss3:*
221: gb_gss4:*
222: gb_gss5:*
223: gb_gss6:*
224: gb_gss7:*
225: gb_gss8:*
226: gb_gss9:*
227: gb_gss10:*
228: gb_gss11:*
229: gb_gss12:*
230: gb_gss13:*
231: gb_gss14:*
232: gb_gss15:*
233: gb_gss16:*
234: gb_gss17:*
235: gb_gss18:*
236: gb_gss19:*
237: gb_gss20:*
238: gb_gss21:*
239: gb_gss22:*
240: gb_gss23:*
241: gb_gss24:*
242: gb_gss25:*
243: gb_gss26:*
244: gb_gss27:*
245: gb_gss28:*
246: gb_gss29:*
247: gb_gss30:*
248: gb_gss31:*
249: gb_gss32:*
250: gb_gss33:*
251: gb_gss34:*
252: em_gss_inv4:*
253: em_gss_rod6:*
254: em_gss_rod7:*
255: em_gss_rod8:*
256: gb_gss35:*
257: gb_gss36:*
258: gb_gss37:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	628.2	30.7	2927	192	AK019504	AK019504 Mus muscu
2	505	24.7	954	144	BF135102	BF135102 601780142
3	497.6	24.3	1079	145	BF141567	BF141567 601788654
4	492	24.1	669	145	BF143546	BF143546 601790896
5	461.8	22.6	1350	192	AK010093	AK010093 Mus muscu
6	460.2	22.5	1347	192	AK010116	AK010116 Mus muscu
7	460.2	22.5	1349	192	AK009300	AK009300 Mus muscu
8	460.2	22.5	1350	192	AK010139	AK010139 Mus muscu
9	458.6	22.4	1340	192	AK010124	AK010124 Mus muscu
10	458.6	22.4	1343	192	AK009479	AK009479 Mus muscu
11	458.6	22.4	1345	192	AK009413	AK009413 Mus muscu
12	458.6	22.4	1345	192	AK009473	AK009473 Mus muscu
13	458.6	22.4	1346	192	AK009359	AK009359 Mus muscu
14	458.6	22.4	1346	192	AK009523	AK009523 Mus muscu
15	458.6	22.4	1346	192	AK009729	AK009729 Mus muscu
16	458.6	22.4	1348	192	AK010058	AK010058 Mus muscu
17	458.6	22.4	1349	192	AK009474	AK009474 Mus muscu
18	458.6	22.4	1349	192	AK009546	AK009546 Mus muscu
19	458.6	22.4	1349	192	AK009571	AK009571 Mus muscu
20	458.6	22.4	1349	192	AK009573	AK009573 Mus muscu
21	458.6	22.4	1349	192	AK010148	AK010148 Mus muscu
22	458.6	22.4	1350	192	AK009459	AK009459 Mus muscu
23	458.6	22.4	1350	192	AK009525	AK009525 Mus muscu
24	458.6	22.4	1350	192	AK009773	AK009773 Mus muscu
25	458.6	22.4	1350	192	AK010061	AK010061 Mus muscu
26	458.6	22.4	1350	192	AK010125	AK010125 Mus muscu
27	458.6	22.4	1350	192	AK010203	AK010203 Mus muscu
28	458.6	22.4	1351	192	AK009428	AK009428 Mus muscu
29	458.6	22.4	1352	192	AK009431	AK009431 Mus muscu
30	458.6	22.4	1356	192	AK010019	AK010019 Mus muscu
31	458.6	22.4	1356	192	AK010035	AK010035 Mus muscu
32	457.6	22.4	1346	192	AK009437	AK009437 Mus muscu
33	457.6	22.4	1347	192	AK010103	AK010103 Mus muscu
34	457.6	22.4	1349	192	AK010026	AK010026 Mus muscu
35	457	22.4	1346	192	AK009560	AK009560 Mus muscu
36	457	22.4	1347	192	AK010236	AK010236 Mus muscu
37	457	22.4	1348	192	AK010231	AK010231 Mus muscu
38	457	22.4	1349	192	AK009875	AK009875 Mus muscu
39	457	22.4	1350	192	AK009537	AK009537 Mus muscu
40	456	22.3	1349	192	AK010106	AK010106 Mus muscu
41	455.4	22.3	1345	192	AK009544	AK009544 Mus muscu
42	455.4	22.3	1349	192	AK009990	AK009990 Mus muscu
43	454	22.2	642	145	BF138035	BF138035 601784732
44	453.8	22.2	1346	192	AK010173	AK010173 Mus muscu
45	450.6	22.0	933	145	BF143673	BF143673 601789739

ALIGNMENTS

RESULT 1
AK019504
LOCUS AK019504 2927 bp mRNA HTC 08-FEB-2001
DEFINITION Mus musculus 0 day neonate skin cDNA, RIKEN full-length enriched library, clone:4632427C23, full insert sequence.
ACCESSION AK019504
VERSION AK019504.1 GI:12859754
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) 0 day neonate skin cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library clone:4632427C23.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (sites)
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Methods Enzymol. 303, 19-44 (1999)
REFERENCE 2 (sites)
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
REFERENCE 3 (sites)
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Kikuchi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Okazaki,Y., Muramatsu,M., Inoue,Y. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
REFERENCE 4 (sites)
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5 (bases 1 to 2927)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Gemomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. cDNA of size comprised between 3 and 7 kb was selected before cloning. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B.
FEATURES Location/Qualifiers
source 1. .2927
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="MGD:MGI:1912763"
/db_xref="MGD:MGI:1926003"
/clone="4632427C23"
/tissue_type="skin"
/clone_lib="RIKEN full-length enriched mouse cDNA library"

```
                     /dev_stage="0 day neonate"
     CDS             153. .662
                     /note="putative"
                     /codon_start=1
                     /protein_id="BAB31766.1"
                     /db_xref="GI:12859755"
                     /translation="MSEILSRVWTVSHRVEIWLLILVAYLLQRNVNSGHLPTKAADPE
                     AFMNVSEIIKHKGYPSEEYEVATEDGYILSVNRIPRGQTRLKKEGSRPVVLLQHGLLG
                     DASNWISNLPNNSLGFILADAGFDVWMGNSRGNTWSRKHKTLSIDQDEFWAFRYKNTL
                     RVEVDTDMP".
     polyA_signal    2908. .2913
                     /note="putative"
     polyA_site      2927
                     /note="putative"
BASE COUNT      837 a    581 c    664 g    845 t
ORIGIN


  Query Match             30.7%;  Score 628.2;  DB 192;  Length 2927;
  Best Local Similarity   84.3%;  Pred. No. 1.3e-149;
  Matches  734;  Conservative    0;  Mismatches  128;  Indels    9;  Gaps    2;

Qy     572 tgagttctgggctttcagttatgatgagatggctaggtttgaccttcctgcagtgataaa 631
           ||  || |    ||  ||||||||||| ||||||||||||||||||||| || ||||||||
Db    1376 TGTCTTTTCCCTTTGTAGTTATGATGAAATGGCTAGGTTTGACCTTCCAGCTGTGATAAA 1435

Qy     632 ctttattttgcagaaaacgggccaggaaaagatctattatgtcggctattcacagggcac 691
           |||||| || |||||||||||||||| ||||| |||||||||||||||| ||||||||||
Db    1436 CTTTATCTTACAGAAAACGGGCCAGAAAAAGGTCTATTATGTCGGCTACTCACAGGGCAC 1495

Qy     692 caccatgggctttattgcattttccaccatgccagagctggctcagaaaatcaaaatgta 751
           |||||||||||||||||||||||||||| |||||||||| ||||| ||||||||||||||
Db    1496 CACCATGGGCTTTATTGCATTTTCCACAATGCCAGAGCTAGCTCATAAAATCAAAATGTA 1555

Qy     752 ttttgctttagcacccatagccactgttaagcatgcaaaaagccccgggaccaaatttttt 811
           |||||| ||||| || |||||||||||||  |||||| ||| || || |||||||| | |
Db    1556 TTTTGCCTTAGCTCCTATAGCCACTGTTAAATATGCAAGAAGTCCTGGTACCAAATTTCT 1615

Qy     812 gttgctgccagatatgatgatcaagggattgtttggcaaaaaagaatttctgtatcagac 871
           | |||||||||||||||||||||||| ||| ||||||| | |||| ||| | || |||||
Db    1616 GCTGCTGCCAGATATGATGATCAAGGTATTATTTGGCAGACAAGAGTTTTTATACCAGAC 1675

Qy     872 cagatttctcagacaacttgttatttacctttgtggccaggtgattcttgatcagatttg 931
           ||||||| ||||||| ||| ||||||||||||||||||||| |||||||||| || || ||
Db    1676 TAGATTTTTCAGACAGCTTTTTATTTACCTTTGTGGCCAGATGATTCTTGACCAAATCTG 1735

Qy     932 tagtaatatcatgttacttctgggtggattcaacaccaacaatatgaacatgagccgagc 991
           || || ||||| ||||| ||||| || || ||||| ||||||||||||||||||| ||||
Db    1736 CAGCAACATCATCTTACTCCTGGGAGGTTTTAACACAAACAATATGAACATGAGCAGAGC 1795

Qy     992 aagtgtatatgctgcccacactcttgctggaacatctgtgcaaaatattctacactggag 1051
           || ||| ||||  ||||| || | ||||||||| |||||||| || ||||| ||||||||
Db    1796 AAATGTGTATGTCGCCCATACACCTGCTGGAACGTCTGTGCAGAACATTCTCCACTGGAG 1855

Qy    1052 ccaggcagtgaattctggtgaactccgggcatttgactgggggagtgagaccaaaaatct 1111
           |||||| |||||||||| ||||| || || |||||||||| ||||||||||||||||||
Db    1856 CCAGGCCGTGAATTCTGGGGAACTTCGTGCCTTTGACTGGGGAAGTGAGACCAAAAATCA 1915

Qy    1112 ggaaaaatgcaatcagccaactcctgtaaggtacagagtcagagatatgacggtccctac 1171
           ||| ||||| |||||||||||||| ||||||||| |||  |||||||| |||||| ||
Db    1916 GGAGAAATGTAATCAGCCAACTCCTATAAGGTACAAAGTTCGAGATATGATGGTCCCAAC 1975

Qy    1172 agcaatgtggacaggaggtcaggactggctttcaaatccagaagacgtgaaaatgctgct 1231
           |||||||||||| |||||||| |||||||||||||||||||| || |||||||   | ||
Db    1976 AGCAATGTGGACTGGAGGTCAAGACTGGCTTTCAAATCCAGATGATGTGAAAACATTACT 2035

Qy    1232 ctctgaggtgaccaacctcatctaccataagaatattcctgaatgggctcacgtggattt 1291
           |||||| || |||||||||||||||||| ||||| |||||||| |||||||| ||||||||
Db    2036 TTCTGAAGTAACCAACCTCATCTACCACAAGAACATTCCTGAGTGGGCTCATGTGGATTT 2095

Qy    1292 catctggggtttggatgctcctcaccgtatgtacaatgaaatcatccatctgatgcagca 1351
           |||||||||  ||||||| ||||| ||| | |||||||||||| || |||||||| |||
Db    2096 CATCTGGGGGCTGGATGCCCCTCAGCGTGTTTACAATGAAATTATACATCTGATGAAGC- 2154

Qy    1352 ggaggagaccaaccttтcccagggacggtgtgaggccgtattgtgaagcatctgacactg 1411
           ||||| |||||||| ||||||||   ||   || | ||||||||||  |||  |
Db    2155 --AGGAGCCCAACCTTCCCCAGGGAACCTGCAGGGTCAAATTGTGAAGCTCCTGCAA--- 2209

Qy    1412 acgatcttaggacaacctcctgagggatggg 1442
              |||||| |||||| | || || |||||
Db    2210 ---GTCTTAGAACAACCCCATGCGGAATGGG 2237


RESULT   2
BF135102
LOCUS       BF135102       954 bp    mRNA            EST       24-OCT-2000
DEFINITION  601780142F1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:4008131 5',
            mRNA sequence.
ACCESSION   BF135102
VERSION     BF135102.1  GI:10974142
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 954)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by:Incyte Genomics, Inc.
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9242  row: c  column: 12
            High quality sequence stop: 695.
FEATURES             Location/Qualifiers
     source          1. .954
                     /organism="Mus musculus"
                     /strain="CZECH II"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4008131"
                     /clone_lib="NCI_CGAP_Lu30"
                     /tissue_type="tumor, metastatic to mammary"
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; transgenic model WNT-1, expression driven by
                     MMTV-LTR enhancer; Cloned unidirectionally.  Primer: Oligo
                     dT.   Library constructed by Life Technologies.
                     Investigator providing samples: Gilbert Smith, NIH"
BASE COUNT      280 a    222 c    222 g    230 t
ORIGIN


  Query Match             24.7%;  Score 505;  DB 144;  Length 954;
  Best Local Similarity   85.3%;  Pred. No. 3e-118;
  Matches  598;  Conservative    0;  Mismatches  100;  Indels    3;  Gaps    3;

Qy     618 cctgcagtgataaacttttattttgcagaaaacgggccaggaaaagatctattatgtcggc 677
           || || |||||||||| || || ||||||||||||||| ||||| ||||||||||||||
Db       1 CCAGCTGTGATAAACTTGATCTTACAGAAAACGGGCCAGAAAAAGGTCTATTATGTCGGC 60

Qy     678 tattcacagggcaccaccatgggctttattgcattttccaccatgccagagctggctcag 737
           || ||||||||||||||||||||||||||||||||||||||||| |||||||||| |||||
Db      61 TACTCACAGGGCACCACCATGGGCTTTATTGCATTTTCCACAATGCCAGAGCTAGCTCAT 120

Qy     738 aaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaaaaagcccc 797
           |||||||||||||||||| ||||| || |||||||||||||  |||||| ||| ||
Db     121 AAAATCAAAATGTATTTTGCCTTAGCTCCTATAGCCACTGTTAAATATGCAAGAAGTCCT 180

Qy     798 gggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggcaaaaaagaa 857
           || |||||||||| || |||||||||||||||||||||||| ||| ||||||| | ||||
```

```
Db    181 GGTACCAAATTTCTGCTGCTGCCAGATATGATGATCAAGGTATTATTTGGCAGACAAGAG 240

Qy    858 tttctgtatcagaccagatttctcagacaacttgttatttacctttgtggccaggtgatt 917
          ||| | || ||||| |||||| |||| || ||| |||||||||| |||||||||| |||||
Db    241 TTTTTATACCAGACTAGATTTTTCAGGCAGCTTTTTATTTACC-TTGTGGCCAGATGATT 299

Qy    918 cttgatcagatttgtagtaatatcatgttacttctgggtggattcaacaccaacaatatg 977
          || || || || || || || ||||| ||||| ||||| || || ||||| |||||||||
Db    300 CTCGACCAAATCTGCAGCAACATCATCTTACTCCTGGGAGGTTTTAACACAAACAATATG 359

Qy    978 aacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctgtgcaaaat 1037
          ||||||||| |||||| ||| |||| ||||| || | |||||||||| |||||||| ||
Db    360 AACATGAGCAGAGCAAATGTGTATGTCGCCCATACACCTGCTGGAACGTCTGTGCAGAAC 419

Qy   1038 attctacactggagccaggcagtgaattctggtgaactccgggcatttgactgggggagt 1097
          ||||| |||||||||||||| |||||||||| ||||| || || |||||||||| |||
Db    420 ATTCTCCACTGGAGCCAGGCCGTGAATTCTGGGGAACTTCGTGCCTTTGACTGGGGAAGT 479

Qy   1098 gagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagagtcagagat 1157
          |||||||||||| ||| ||||| |||||||||||||| ||||||||| |||  |||||
Db    480 GAGACCAAAAATCAGGAGAAATGTAATCAGCCAACTCCTATAAGGTACAAAGTTCGAGAT 539

Qy   1158 atgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatccagaagac 1217
          |||| |||||| |||||||||||||| |||||||| |||||| ||||||||||||| ||
Db    540 ATGATGGTCCCAACAGCAATGTGGACTGGAGGTCAAGACTGG-TTTCAAATCCAGATGAT 598

Qy   1218 gtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattcctgaatgg 1277
          ||||||    | || ||||| || ||||| |||||||||| ||||| ||||||||   |
Db    599 GTGAAACAATTACTTTCTGAAGTAACCAAACTCATCTACCACAAGAACATTCCTGAGGTG 658

Qy   1278 gctcacgtggatttcatctggggtttggatgctcctcaccg 1318
          | ||| ||||| |||||||||||  |||||| ||||| ||
Db    659 GGTCATGTGGA-TTCATCTGGGGGGTGGATGCCCCTCAGCG 698


RESULT   3
BF141567
LOCUS       BF141567     1079 bp    mRNA            EST       24-OCT-2000
DEFINITION  601788654F1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:4016413 5',
            mRNA sequence.
ACCESSION   BF141567
VERSION     BF141567.1  GI:10980607
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1079)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by:Incyte Genomics, Inc.
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9263  row: 1  column: 14
            High quality sequence stop: 663.
FEATURES             Location/Qualifiers
     source          1. .1079
                     /organism="Mus musculus"
                     /strain="CZECH II"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4016413"
                     /clone_lib="NCI_CGAP_Lu30"
                     /tissue_type="tumor, metastatic to mammary"
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; transgenic model WNT-1, expression driven by
                     MMTV-LTR enhancer; Cloned unidirectionally.  Primer: Oligo
                     dT.   Library constructed by Life Technologies.
                     Investigator providing samples: Gilbert Smith, NIH"
BASE COUNT      312 a    267 c    263 g    237 t
ORIGIN


  Query Match           24.3%;  Score 497.6;  DB 145;  Length 1079;
  Best Local Similarity   84.1%;  Pred. No. 2.4e-116;
  Matches  621;  Conservative    0;  Mismatches  109;  Indels    8;  Gaps    5;

Qy    381 aagaagacaggttccaggcctgtggtgttactgcagcatggcctagttggaggtgctagc 440
          |||||| ||| |||||||| |||||||||||||| ||||| ||  |||| | |||||||
Db     20 AAGAAGGAAGGATCCAGGCCAGTGGTGTTACTGCAACATGGTCTTCTTGGGGATGCTAGC 79

Qy    441 aactggatttccaacctgcccaacaatagcctgggcttcattctggcagatgctggtttt 500
          |||||||||||||||||||||||||| |||||||| || ||||||||||||| ||||||
Db     80 AACTGGATTTCCAACCTGCCCAACAACAGCCTGGGTTTTATTCTGGCAGATGCAGGTTTT 139

Qy    501 gacgtgtggatggggaacagcaggggaaacgcc-tggtctcgaaaacacaagacactctc 559
          || |||||||||| |||||||| |||||| || |||||||| || |||||||| |||||
Db    140 GATGTGTGGATGGGAAACAGCAGAGGAAACACCGTGGTCTCGGAAGCACAAGACCCTCTC 199

Qy    560 catagaccaagatgagttctgggctttcagttatgatgagatggctaggtttgaccttcc 619
          |||||| ||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    200 CATAGATCAAGATGAGTTCTGGGCTTTCAGTTATGATGAAATGGCTAGGTTTGACCTTCC 259

Qy    620 tgcagtgataaactttattttgcagaaaacgggccaggaaaagatctattatgtcggcta 679
          || |||||||||||||| || |||||||||||||| ||||| ||||||||||||||||
Db    260 AGCTGTGATAAACTTTATCTTACAGAAAACGGGCCAGAAAAAGGTCTATTATGTCGGCTA 319

Qy    680 ttcacagggcaccaccatgggctttattgcattttccaccatgccagagctggctcagaa 739
          |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db    320 CTCACAGGGCACCACCATGGGCTTTATTGCATTTTCCACAATGCCAGAGCTAGCTCATAA 379

Qy    740 aatcaaaaatgtattttgctttagcacccatagccactgttaagcatgcaaaaagccccgg 799
          |||||||||||||||| || ||||| || |||||||||||||  |||||| ||| || ||
Db    380 AATCAAAATGTATTTGGCCTTAGCTCCTATAGCCACTGTTAAATATGCAAGAAGTCCTGG 439

Qy    800 gaccaaatttttgttgctgccagatatgatgatcaagggattgtttggcaaaaaagaatt 859
          |||||||| || |||||||||||||||||||||||| ||| ||||||| | |||| ||
Db    440 TACCAAATTTCTGCTGCTGCCAGATATGATGATCAAGGTATTATTTGGCAGACAAGAGTT 499

Qy    860 tctgtatcagaccagatttctcagacaacttgttatttacctttgtggccaggtgattct 919
          | | || ||||| |||||| |||| || ||| ||||||||||||||||||| ||| |||||||
Db    500 TTTATACCAGACTAGATTTTTCAGGCAGCTTTTTATTTACCTTTGTGGGCAGATGATTCT 559

Qy    920 tgatcagatttgtagtaatatcatgttacttctgggtggattcaacaccaacaatatgaa 979
          || || || || || || ||||| ||||| ||||| ||.|| ||||| ||| |||||||
Db    560 CGACCAAATCTGCAGCAACATCATCTTACTCCTGGGAGGTTTTAACACAAACCATATGAA 619

Qy    980 catgagccgagca--agtgtatatgctgcccacactcttgctggaacatctgtg---caa 1034
          ||||||| ||||| |     | ||  ||||| |  | |||||||||| ||| |   | |
Db    620 CATGAGCAGAGCAACATGTGCTCTGTCGCCCATAAACCTGCTGGAACGTCTTGGTGCCGA 679

Qy   1035 aatattctacactggagccagg-cagtgaattctggtg-aactccgggcatttgactggg 1092
          |  ||||| |||||||||||| | |||||||||| | | || || || |||||||| |
Db    680 ACCATTCTCCACTGGAGCCAGGCCGGTGAATTCTGGGGCACCTTCGTGCCTTTGACTGTG 739

Qy   1093 ggagtgagaccaaaaatc 1110
          ||   | ||| |  || |
Db    740 GGCAAGCGACAACCAAGC 757


RESULT   4
BF143546
LOCUS       BF143546      669 bp    mRNA            EST       24-OCT-2000
DEFINITION  601790896F1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:4021870 5',
            mRNA sequence.
ACCESSION   BF143546
VERSION     BF143546.1  GI:10982586
KEYWORDS    EST.
SOURCE      house mouse.
```

```
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 669)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by:Incyte Genomics, Inc.
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9277  row: o  column: 23
            High quality sequence stop: 659.
FEATURES             Location/Qualifiers
     source          1. .669
                     /organism="Mus musculus"
                     /strain="CZECH II"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4021870"
                     /clone_lib="NCI_CGAP_Lu30"
                     /tissue_type="tumor, metastatic to mammary"
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; transgenic model WNT-1, expression driven by
                     MMTV-LTR enhancer; Cloned unidirectionally.  Primer: Oligo
                     dT.   Library constructed by Life Technologies.
                     Investigator providing samples: Gilbert Smith, NIH"
BASE COUNT      195 a    140 c    165 g    169 t
ORIGIN


  Query Match           24.1%;  Score 492;  DB 145;  Length 669;
  Best Local Similarity   85.5%;  Pred. No. 5.9e-115;
  Matches  571;  Conservative    0;  Mismatches   95;  Indels    2;  Gaps    2;

Qy      82 tagttgttacattggcaggaa-aaaataaatgcagatgttggaccatgttggaaaccttg 140
           ||||||||||||    || ||  |||  |||||| ||||||||||||||||  |||| ||||
Db       3 TAGTTGTTACATCCTTAGCAACAAAGCAAATGCCGATGTTGGACCATGTCAGAAATCTTG 62

Qy     141 tcaagacagtggattgtctcacacagaatggaaatgtggcttctgattctggtggcgtat 200
           ||||||  |||||  ||| || |||||| |||| || ||||||||||||||||| || |||
Db      63 TCAAGAGTGTGGACTGTTTCGCACAGAGTGGAGATATGGCTTCTGATTCTGGTAGCATAT 122

Qy     201 atgttccagagaaatgtgaattcagtacatatgccaactaaagctgtggacccagaagca 260
           |  |||| |||||||||||| || | |||| |||| || ||||||| ||| |||||||||
Db     123 TTACTCCAAAGAAATGTGAACTCGGGACATTTGCCCACGAAAGCTGCGGATCCAGAAGCA 182

Qy     261 ttcatgaatattagtgaaatcatccaacatcaaggctatccctgtgaggaatatgaagtc 320
           |||||||||  |||| |||||||||  ||||   | || ||||||  |||||||  ||||||||
Db     183 TTCATGAATGTTAGCGAAATCATCAAACACAAGGGTTATCCCAGTGAGGAGTATGAAGTT 242

Qy     321 gcaactgaagatgggtatatccttttctgttaacaggattcctcgaggcctagtgcaacct 380
           |||||  ||||||||||| ||||||||||| |||||  || |||||  || |     |
Db     243 GCAACAGAAGATGGGTACATCCTTTCTGTGAACAGAATCCCTCGGGGACAGACACGGTTA 302

Qy     381 aagaagacaggttccaggcctgtggtgttactgcagcatggcctagttggaggtgctagc 440
           ||||||  ||| |||||||| ||||||||||||||| |||||  ||  ||||  | |||||||
Db     303 AAGAAGGAAGGATCCAGGCCAGTGGTGTTACTGCAACATGGTCTTCTTGGGGATGCTAGC 362

Qy     441 aactggatttccaacctgcccaacaatagcctgggcttcattctggcagatgctggtttt 500
           ||||||||||||||||||||||||||||| |||||||  || ||||||||||||  || |||
Db     363 AACTGGATTTCCAACCTGCCCAACAACAGCCTGGGTTTTATTCTGGCAGATGCAGG-TTT 421

Qy     501 gacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacacaagacactctcc 560
           ||  |||||||||| |||||||| |||||| ||||||||| || |||||||| ||||||
Db     422 GATGTGTGGATGGGAAACAGCAGAGGAAACACCTGGTCTCGGAAGCACAAGACCCTCTCC 481

Qy     561 atagaccaagatgagttctgggctttcagttatgatgagatggctaggtttgaccttcct 620
           |||||  |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     482 ATAGATCAAGATGAGTTCTGGGCTTTCAGTTATGATGAAATGGCTAGGTTTGACCTTCCA 541

Qy     621 gcagtgataaactttattttgcagaaaacgggccaggaaaagatctattatgtcggctat 680
           || |||||||||||||| || ||||||||||||||| ||||| ||||||||||||||||
Db     542 GCTGTGATAAACTTTATCTTACAGAAAACGGGCCAGAAAAAGGTCTATTATGTCGGCTAC 601

Qy     681 tcacagggcaccaccatgggctttattgcattttccaccatgccagagctggctcagaaa 740
           |||||||||||||||||||||||||||||||||||||||||| |||||||||| ||||| ||
Db     602 TCACAGGGCACCACCATGGGCTTTATTGCATTTTCCACAATGCCAGAGCTAGCTCATAAC 661

Qy     741 atcaaaat 748
           |||  |||
Db     662 ATCCAAAT 669


RESULT   5
AK010093
LOCUS       AK010093     1350 bp    mRNA            HTC       08-FEB-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310067K20, full insert sequence.
ACCESSION   AK010093
VERSION     AK010093.1  GI:12845297
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310067K20.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods Enzymol. 303, 19-44 (1999)
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Kikuchi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T.,
            Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T.,
            Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Okazaki,Y., Muramatsu,M., Inoue,Y. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1350)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
```

Exploration Research Group, RIKEN Gemomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)

COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 25.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI. Host: SOLR.

FEATURES Location/Qualifiers

source 1. .1350
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="MGD:MGI:1898392"
/db_xref="MGD:MGI:1914967"
/clone="2310067K20"
/sex="male"
/tissue_type="tongue"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"

CDS 53. .1240
/note="putative"
/codon_start=1
/protein_id="BAB26697.1"
/db_xref="GI:12845298"
/translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLVASATNWITNLPNNSLA FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM AED"

BASE COUNT 396 a 290 c 286 g 378 t

ORIGIN

Query Match 22.6%; Score 461.8; DB 192; Length 1350;
Best Local Similarity 63.8%; Pred. No. 3.6e-107;
Matches 700; Conservative 0; Mismatches 397; Indels 0; Gaps 0;

```
Qy    250 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 309
          |||| ||||||  ||||||| ||||| | || ||     |    || |||||  ||||||
Db    132 ACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTACTGGGGATATCCAAGTGAGG 191

Qy    310 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 369
          ||||||||||  ||||||||||| || || ||    ||  | || ||||||   ||
Db    192 AATATGAAGTTGTTACTGAAGATGGCTACATTCTGGGGGTCTATAGAATTCCTTATGGGA 251

Qy    370 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 429
              |  || |||| |  ||    || |||||||  |   ||||||||||  | ||||
Db    252 AGAAAAATTCTGAGAATATCGGCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGGTTG 311

Qy    430 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 489
            |   || |  |||||||| | || ||||| ||||| ||||||| ||||||||| ||||
Db    312 CATCAGCCACAAACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAG 371

Qy    490 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 549
          |||||||  | ||| ||||||  |||||||||  | || ||  | ||||| || ||| |
Db    372 ATGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCCGGAAAAATG 431
```

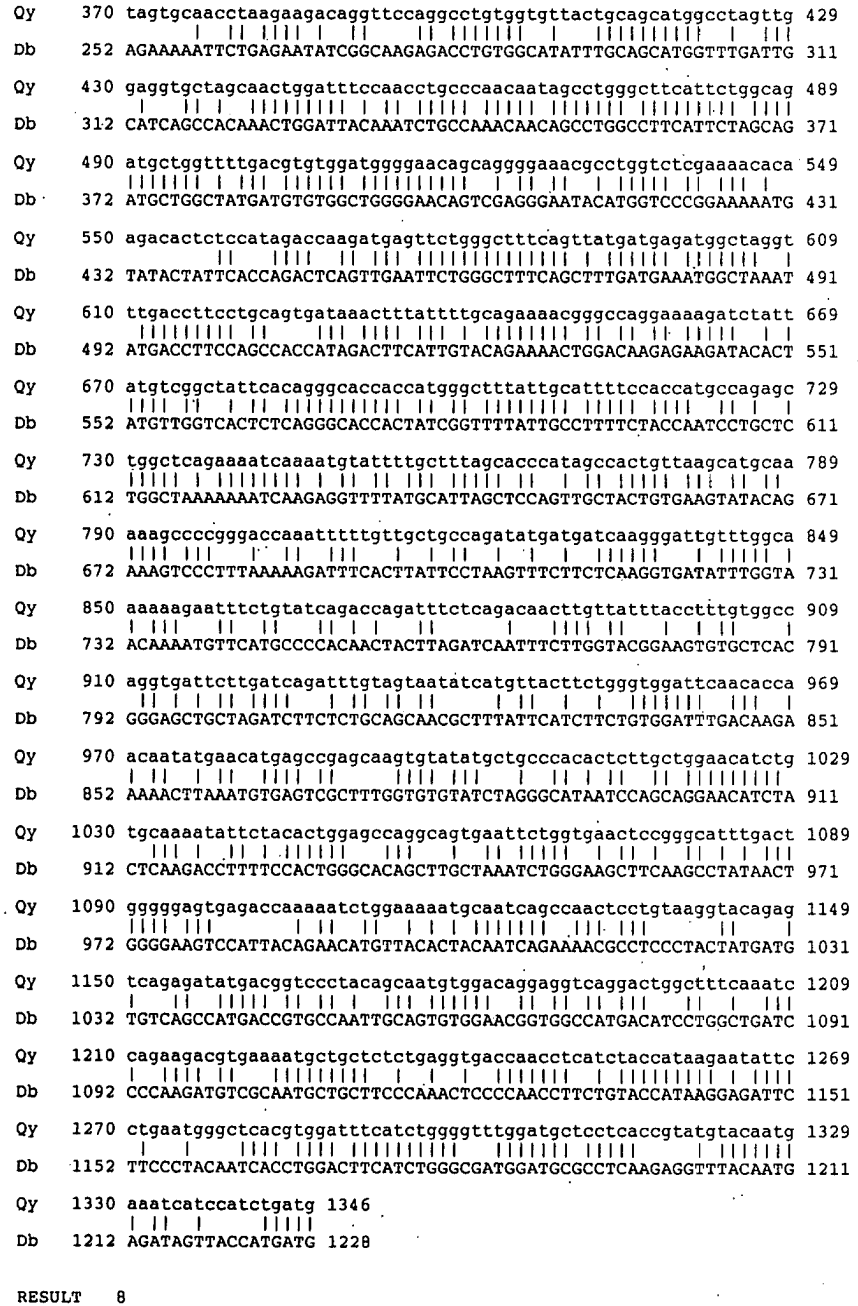
```
Qy    550 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 609
                  ||   ||||  || ||| ||||||||||||||| | |||||| |||||||  |
Db    432 TATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGGCTAAAT 491

Qy    610 ttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 669
          |||||||| ||    ||| |||| ||| | |||||||| || || || |||||  | |
Db    492 ATGACCTTCCAGCCACCATAGACTTCATTGTACAGAAAACTGGACAAGAGAAGATACACT 551

Qy    670 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 729
          |||| ||  | || ||||||||||| || || ||||||| ||||| |||| || |  |
Db    552 ATGTTGGTCACTCTCAGGGCACCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTC 611

Qy    730 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 789
          ||||| | |||||||| | || || ||| ||||| ||  | || ||||| ||| || ||
Db    612 TGGCTAAAAAAATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAG 671

Qy    790 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 849
          |||| |||   |  ||  |||   | | ||  |  |  |  ||||||   | ||||| |
Db    672 AAAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGATATTTGGTA 731

Qy    850 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 909
          | |||   ||  ||   || | |   ||       |   |||| ||   |  | ||   |
Db    732 ACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTTGGTACGGAAGTGTGCTCAC 791

Qy    910 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 969
          || | | || ||||   | || || ||     | ||  |. |  |||||||  |||  |
Db    792 GGGAGCTGCTAGATCTTCTCTGCAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGA 851

Qy    970 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 1029
          | ||  | ||  |||| ||      ||| |||   |  || | ||  || |||||||||
Db    852 AAAACTTAAATGTGAGTCGCTTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTA 911

Qy    1030 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 1089
           ||| |  || | ||||||  |||   |  || |||||  | || |  || | | |||
Db    912 CTCAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGCCTATAACT 971

Qy    1090 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1149
           ||||||||       | ||  ||  | | | |||||||  ||| |||      ||   |
Db    972 GGGGGAGTCCATTACAGAACATGTTACACTACAATCAGAAAACGCCTCCCTACTATGATG 1031

Qy    1150 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1209
           |  ||  ||||| || || |  ||| ||||||  || || || |||   || |  |||
Db    1032 TGTCAGCCATGACCGTGCCAATTGCAGTGTGGAACGGTGGCCATGACATCCTGGCTGATC 1091

Qy    1210 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1269
           |  |||| ||   |||||||||  |  |  | |||||||  | |||||||||  | ||||
Db    1092 CCCAAGATGTCGCAATGCTGCTTCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTC 1151

Qy    1270 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1329
           |   |    |||| |||| ||||||||||   |||||||  |||||     | |||||||
Db    1152 TTCCCTACAATCACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATG 1211

Qy    1330 aaatcatccatctgatg 1346
           | ||  |     |||||
Db    1212 AGATAGTTACCATGATG 1228
```

RESULT 6
AK010116
LOCUS AK010116 1347 bp mRNA HTC 08-FEB-2001
DEFINITION Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310068O05, full insert sequence.
ACCESSION AK010116
VERSION AK010116.1 GI:12845334
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library clone:2310068O05.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (sites)

```
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods Enzymol. 303, 19-44 (1999)
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Kikuchi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T.,
            Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T.,
            Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Okazaki,Y., Muramatsu,M., Inoue,Y. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1347)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 25.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES            Location/Qualifiers
     source         1. .1347
                    /organism="Mus musculus"
                    /strain="C57BL/6J"
                    /db_xref="taxon:10090"
                    /db_xref="MGD:MGI:1895801"
                    /db_xref="MGD:MGI:1914967"
                    /clone="2310068005"
                    /sex="male"
                    /tissue_type="tongue"
                    /clone_lib="RIKEN full-length enriched mouse cDNA library"
                    /dev_stage="adult"
     CDS            51. .1238
                    /note="putative"
                    /codon_start=1
                    /protein_id="BAB26711.1"
                    /db_xref="GI:12845335"
                    /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                    SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
                    FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                    GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
                    FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFGV
                    YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                    AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                    AED"
BASE COUNT      393 a    290 c    284 g    380 t
ORIGIN
```

```
  Query Match             22.5%;  Score 460.2;  DB 192;  Length 1347;
  Best Local Similarity   63.7%;  Pred. No. 9.3e-107;
  Matches  699;  Conservative    0;  Mismatches  398;  Indels    0;  Gaps    0;

Qy    250 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 309
          |||| |||||| ||||||| ||||| | || ||     |    || |||||  ||||||
Db    130 ACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTACTGGGGATATCCAAGTGAGG 189

Qy    310 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 369
          |||| ||||| |  |||||||||||| || || ||    ||  | || ||||||   ||
Db    190 AATATGAAGTTGTTACTGAAGATGGCTACATTCTGGGGGTCTATAGAATTCCTTATGGGA 249

Qy    370 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 429
              || |||| |  ||  || |||||| |  |||||||||  |  |||
Db    250 AGAAAAATTCTGAGAATATCGGCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTG 309

Qy    430 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 489
          |  || |  |||||||||| | || ||||| ||||| ||||||| |||||||||  ||||
Db    310 CATCAGCCACAAACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAG 369

Qy    490 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 549
          ||||||| | ||| |||||| ||||||||||  | || ||  | ||||| || ||| |
Db    370 ATGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCCGGAAAAATG 429

Qy    550 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 609
                ||   ||||  || ||| |||||||||||||| | |||||| |||||||  |
Db    430 TATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGGCTAAAT 489

Qy    610 ttgaccttcctgcagtgataaactttatttttgcagaaaacgggccaggaaaagatctatt 669
          |||||||||  ||    ||| |||| ||| | |||||||| || || || ||||  | |
Db    490 ATGACCTTCCAGCCACCATAGACTTCATTGTACAGAAAACTGGACAAGAGAAGATACACT 549

Qy    670 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 729
          |||| ||  | || |||||||||||| || || |||||||| |||| ||||  || |  |
Db    550 ATGTTGGTCACTCTCAGGGCACCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTC 609

Qy    730 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 789
          ||||| | |||||||| | || || ||| ||||| ||  | || ||||| ||| || ||
Db    610 TGGCTAAAAAAATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAG 669

Qy    790 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 849
          |||| |||   |  ||  |||   | |  |  |  |  ||||||  | |||||  |
Db    670 AAAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGATATTTGGTA 729

Qy    850 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccttttgtggcc 909
          | |||   ||  ||  || | |   ||       |   |||| ||   |  | ||    |
Db    730 ACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTTGGTACGGAAGTGTGCTCAC 789

Qy    910 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 969
          || | | || ||||   | || || ||     | ||  |  |  |||||||  |||  |
Db    790 GGGAGCTGCTAGATCTTCTCTGCAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGA 849

Qy    970 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 1029
          | ||  | ||  |||| ||    |||| |||   |  || | ||  || ||||||||||
```

```
Db    850 AAAACTTAAATGTGAGTCGCTTTGGTGTGTATCTAGGGCATAATCCAGCAGGAACATCTA 909

Qy   1030 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 1089
          ||| |  || | ||||||   |||   |  || |||||  | || |  || | | |||
Db    910 CTCAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGCCTATAACT 969

Qy   1090 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1149
          |||| |||       | ||  ||  | | | |||||||  ||| |||      ||    |
Db    970 GGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAACGCCTCCCTACTATGATG 1029

Qy   1150 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1209
          |   ||  ||||| || || |  ||| ||||||  || || || |||   ||  |  |||
Db   1030 TGTCAGCCATGACCGTGCCAATTGCAGTGTGGAACGGTGGCCATGACATCCTGGCTGATC 1089

Qy   1210 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1269
          |  |||| ||   |||||||||  |  |  |  |||||||  | |||||||| | ||||
Db   1090 CCCAAGATGTCGCAATGCTGCTTCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTC 1149

Qy   1270 ctgaatgggctcacgtggatttcatctgggtttggatgctcctcaccgtatgtacaatg 1329
          |   |    |||| |||| ||||||||||   |||||| |||||    | |||||||
Db   1150 TTCCCTACAATCACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATG 1209

Qy   1330 aaatcatccatctgatg 1346
          | ||  |    |||||
Db   1210 AGATAGTTACCATGATG 1226


RESULT   7
AK009300
LOCUS       AK009300     1349 bp    mRNA            HTC       08-FEB-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310011G18, full insert sequence.
ACCESSION   AK009300
VERSION     AK009300.1  GI:12844011
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310011G18.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods Enzymol. 303, 19-44 (1999)
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Kikuchi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T.,
            Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T.,
            Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Okazaki,Y., Muramatsu,M., Inoue,Y. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1349)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 25.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES              Location/Qualifiers
     source           1. .1349
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /db_xref="MGD:MGI:1895751"
                      /db_xref="MGD:MGI:1914967"
                      /clone="2310011G18"
                      /sex="male"
                      /tissue_type="tongue"
                      /clone_lib="RIKEN full-length enriched mouse cDNA library"
                      /dev_stage="adult"
     CDS              53. .1240
                      /note="putative"
                      /codon_start=1
                      /protein_id="BAB26201.1"
                      /db_xref="GI:12844012"
                      /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                      SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
                      FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                      GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
                      FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFGV
                      YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                      AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                      AED"
BASE COUNT      395 a    290 c    284 g    380 t
ORIGIN


  Query Match             22.5%;  Score 460.2;  DB 192;  Length 1349;
  Best Local Similarity   63.7%;  Pred. No. 9.3e-107;
  Matches  699;  Conservative    0;  Mismatches  398;  Indels    0;  Gaps    0;

Qy    250 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 309
          |||| |||||| |||||| ||||| | || ||     |   || |||||  ||||||
Db    132 ACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTACTGGGGATATCCAAGTGAGG 191

Qy    310 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 369
          ||| |||||| |  |||||||||||| || || ||    ||  | || ||||||   ||
Db    192 AATATGAAGTTGTTACTGAAGATGGCTACATTCTGGGGGTCTATAGAATTCCTTATGGGA 251
```



```
Qy    370 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 429
Db    252 AGAAAAATTCTGAGAATATCGGCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTG 311

Qy    430 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 489
Db    312 CATCAGCCACAAACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAG 371

Qy    490 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 549
Db    372 ATGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCCGGAAAAATG 431

Qy    550 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 609
Db    432 TATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGGCTAAAT 491

Qy    610 ttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 669
Db    492 ATGACCTTCCAGCCACCATAGACTTCATTGTACAGAAAACTGGACAAGAGAAGATACACT 551

Qy    670 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 729
Db    552 ATGTTGGTCACTCTCAGGGCACCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTC 611

Qy    730 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 789
Db    612 TGGCTAAAAAAATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAG 671

Qy    790 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 849
Db    672 AAAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGATATTTGGTA 731

Qy    850 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 909
Db    732 ACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTTGGTACGGAAGTGTGCTCAC 791

Qy    910 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 969
Db    792 GGGAGCTGCTAGATCTTCTCTGCAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGA 851

Qy    970 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 1029
Db    852 AAAACTTAAATGTGAGTCGCTTTGGTGTGTATCTAGGGCATAATCCAGCAGGAACATCTA 911

Qy   1030 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 1089
Db    912 CTCAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGCCTATAACT 971

Qy   1090 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1149
Db    972 GGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAACGCCTCCCTACTATGATG 1031

Qy   1150 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1209
Db   1032 TGTCAGCCATGACCGTGCCAATTGCAGTGTGGAACGGTGGCCATGACATCCTGGCTGATC 1091

Qy   1210 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1269
Db   1092 CCCAAGATGTCGCAATGCTGCTTCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTC 1151

Qy   1270 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1329
Db   1152 TTCCCTACAATCACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATG 1211

Qy   1330 aaatcatccatctgatg 1346
Db   1212 AGATAGTTACCATGATG 1228
```

RESULT 8
AK010139
LOCUS AK010139 1350 bp mRNA HTC 08-FEB-2001
DEFINITION Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310069P19, full insert sequence.
ACCESSION AK010139
VERSION AK010139.1 GI:12845371
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library clone:2310069P19.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (sites)
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Methods Enzymol. 303, 19-44 (1999)
REFERENCE 2 (sites)
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
REFERENCE 3 (sites)
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Kikuchi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Okazaki,Y., Muramatsu,M., Inoue,Y. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
REFERENCE 4 (sites)
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5 (bases 1 to 1350)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Gemomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 25.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.

```
          Host: SOLR.
FEATURES             Location/Qualifiers
     source          1. .1350
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /db_xref="MGD:MGI:1898377"
                     /db_xref="MGD:MGI:1914967"
                     /clone="2310069P19"
                     /sex="male"
                     /tissue_type="tongue"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             53. .1240
                     /note="putative"
                     /codon_start=1
                     /protein_id="BAB26725.1"
                     /db_xref="GI:12845372"
                     /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                     SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
                     FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                     GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPIATVKYTESPFKKISLIPK
                     FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV
                     YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                     AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                     AED"
BASE COUNT      398 a    289 c    283 g    380 t
ORIGIN
```

```
  Query Match             22.5%;  Score 460.2;  DB 192;  Length 1350;
  Best Local Similarity   63.7%;  Pred. No. 9.3e-107;
  Matches  699;  Conservative    0;  Mismatches  398;  Indels    0;  Gaps    0;

Qy     250 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 309
           |||| ||||||  |||||||| ||||| | || ||     |    || |||||  ||||||
Db     132 ACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTACTGGGGATATCCAAGTGAGG 191

Qy     310 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 369
           ||||||||||  |  ||||||||||| || || ||    ||  | || ||||||   ||
Db     192 AATATGAAGTTGTTACTGAAGATGGCTACATTCTGGGGGTCTATAGAATTCCTTATGGGA 251

Qy     370 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 429
               |  || |||| |  ||    || |||||||  |   ||||||||||  |  |||
Db     252 AGAAAAATTCTGAGAATATCGGCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTG 311

Qy     430 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 489
            |   || |  |||||||||  | || ||||| ||||| |||||| |||||||| ||||
Db     312 CATCAGCCACAAACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAG 371

Qy     490 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 549
           |||||.|| | ||| |||||| |||||||||  | || ||. | ||||| || ||| |
Db     372 ATGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCCGGAAAAATG 431

Qy     550 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 609
                 ||   ||||  || ||| |||||||||||||| | |||||| |||||||  |
Db     432 TATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGGCTAAAT 491

Qy     610 ttgaccttcctgcagtgataaactttatttttgcagaaaacgggccaggaaaagatctatt 669
           |||||||||  ||    ||| |||| ||| | |||||||| || || || |||||  | |
Db     492 ATGACCTTCCAGCCACCATAGACTTCATTGTACAGAAAACTGGACAAGAGAAGATACACT 551

Qy     670 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 729
           |||| ||  |  || |||||||||||| || || |||||||| ||||| ||||  || |  |
Db     552 ATGTTGGTCACTCTCAGGGCACCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTC 611

Qy     730 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 789
           ||||| | |||||||| | || || ||| ||||| || || || ||||| ||| || ||
Db     612 TGGCTAAAAAAATCAAGAGGTTTTATGCATTAGCTCCAATTGCTACTGTGAAGTATACAG 671

Qy     790 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 849
           |||| |||   |  ||  |||  |  | || |  |  |  ||||||  | ||||| |
Db     672 AAAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGATATTTGGTA 731

Qy     850 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 909
           | |||   ||  ||   || | |   ||       |  |||| ||   |  | ||    |
Db     732 ACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTTGGTACGGAAGTGTGCTCAC 791

Qy     910 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 969
           || | | || |||| . | || || ||     | ||  |  |  |||||||  |||  |
Db     792 GGGAGCTGCTAGATCTTCTCTGCAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGA 851

Qy     970 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 1029
           | ||  | ||  |||| ||      ||| |||   |  || | ||  || ||||||||||
Db     852 AAAACTTAAATGTGAGTCGCTTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTA 911

Qy    1030 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 1089
           |||  |  || | ||||||   |||   |  || |||||  | || |  || | | |||
Db     912 CTCAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGCCTATAACT 971

Qy    1090 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1149
           ||||  |||       | ||  ||  | | | |||||||  |||  |||      ||   |
Db     972 GGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAACGCCTCCCTACTATGATG 1031

Qy    1150 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1209
           |   ||  ||||| || || |  ||| |||||| .|| || || |||   ||  |  |||
Db    1032 TGTCAGCCATGACCGTGCCAATTGCAGTGTGGAACGGTGGCCATGACATCCTGGCTGATC 1091

Qy    1210 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1269
           |  |||| ||   |||||||| |  |  | |||||||  | |||||||| | ||||
Db    1092 CCCAAGATGTCGCAATGCTGCTTCCCAAACTCCCCAACCTTTTGTACCATAAGGAGATTC 1151

Qy    1270 ctgaatgggctcacgtggatttcatctggggttttggatgctcctcaccgtatgtacaatg 1329
           |   |    |||| |||| ||||||||||   |||||||  |||||      | |||||||
Db    1152 TTCCCTACAATCACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATG 1211

Qy    1330 aaatcatccatctgatg 1346
           | ||  |     |||||
Db    1212 AGATAGTTACCATGATG 1228
```

```
RESULT   9
AK010124
LOCUS       AK010124     1340 bp    mRNA            HTC       08-FEB-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310069D09, full insert sequence.
ACCESSION   AK010124
VERSION     AK010124.1  GI:12845346
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310069D09.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods Enzymol. 303, 19-44 (1999)
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Kikuchi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T.,
            Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T.,
            Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Okazaki,Y., Muramatsu,M., Inoue,Y. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
```

```
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1340)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 25.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES              Location/Qualifiers
     source           1. .1340
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /db_xref="MGD:MGI:1895811"
                      /db_xref="MGD:MGI:1914967"
                      /clone="2310069D09"
                      /sex="male"
                      /tissue_type="tongue"
                      /clone_lib="RIKEN full-length enriched mouse cDNA library"
                      /dev_stage="adult"
     CDS              47. .1234
                      /note="putative"
                      /codon_start=1
                      /protein_id="BAB26715.1"
                      /db_xref="GI:12845347"
                      /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                      SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
                      FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                      GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
                      FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV
                      YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                      AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                      AED"
BASE COUNT      393 a    287 c    282 g    378 t
ORIGIN
```

Query Match 22.4%; Score 458.6; DB 192; Length 1340;

Best Local Similarity 63.6%; Pred. No. 2.4e-106;
Matches 698; Conservative 0; Mismatches 399; Indels 0; Gaps 0;

```
Qy    250 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 309
Db    126 ACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTACTGGGGATATCCAAGTGAGG 185

Qy    310 aatatgaagtcgcaactgaagatgggtatatccctttctgttaacaggattcctcgaggcc 369
Db    186 AATATGAAGTTGTTACTGAAGATGGCTACATTCTGGGGGTCTATAGAATTCCTTATGGGA 245

Qy    370 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 429
Db    246 AGAAAAATTCTGAGAATATCGGCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTG 305

Qy    430 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 489
Db    306 CATCAGCCACAAACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAG 365

Qy    490 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 549
Db    366 ATGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCCGGAAAAATG 425

Qy    550 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 609
Db    426 TATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGGCTAAAT 485

Qy    610 ttgaccttcctgcagtgataaacttttattttgcagaaaacgggccaggaaaagatctatt 669
Db    486 ATGACCTTCCAGCCACCATAGACTTCATTGTACAGAAAACTGGACAAGAGAAGATACACT 545

Qy    670 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 729
Db    546 ATGTTGGTCACTCTCAGGGCACCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTC 605

Qy    730 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 789
Db    606 TGGCTAAAAAAATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAG 665

Qy    790 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 849
Db    666 AAAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGATATTTGGTA 725

Qy    850 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 909
Db    726 ACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTTGGTACGGAAGTGTGCTCAC 785

Qy    910 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 969
Db    786 GGGAGCTGCTAGATCTTCTCTGCAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGA 845

Qy    970 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 1029
Db    846 AAAACTTAAATGTGAGTCGCTTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTA 905

Qy   1030 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 1089
Db    906 CTCAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGCCTATAACT 965

Qy   1090 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1149
Db    966 GGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAACGCCTCCCTACTATGATG 1025

Qy   1150 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1209
Db   1026 TGTCAGCCATGACCGTGCCAATTGCAGTGTGGAACGGTGGCCATGACATCCTGGCTGATC 1085

Qy   1210 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1269
Db   1086 CCCAAGATGTCGCAATGCTGCTTCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTC 1145

Qy   1270 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1329
```

```
Db   1146 TTCCCTACAATCACCTGGACTTCATCTGGGCGATGGATGCGCCTCAGGAGGTTTACAATG 1205

Qy   1330 aaatcatccatctgatg 1346
          | ||  |    |||||
Db   1206 AGATAGTTACCATGATG 1222


RESULT  10
AK009479
LOCUS       AK009479     1343 bp    mRNA            HTC       08-FEB-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310022K08, full insert sequence.
ACCESSION   AK009479
VERSION     AK009479.1  GI:12844303
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310022K08.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods Enzymol. 303, 19-44 (1999)
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Kikuchi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T.,
            Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T.,
            Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Okazaki,Y., Muramatsu,M., Inoue,Y. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1343)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 25.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES             Location/Qualifiers
     source          1. .1343
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /db_xref="MGD:MGI:1896992"
                     /db_xref="MGD:MGI:1914967"
                     /clone="2310022K08"
                     /sex="male"
                     /tissue_type="tongue"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             50. .1237
                     /note="putative"
                     /codon_start=1
                     /protein_id="BAB26316.1"
                     /db_xref="GI:12844304"
                     /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                     SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
                     FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                     GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
                     FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV
                     YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                     AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                     AED"
BASE COUNT      394 a    289 c    282 g    378 t
ORIGIN
```

Query Match 22.4%; Score 458.6; DB 192; Length 1343;
Best Local Similarity 63.6%; Pred. No. 2.4e-106;
Matches 698; Conservative 0; Mismatches 399; Indels 0; Gaps 0;

```
Qy    250 accagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 309
          |||| ||||||  ||||||| |||||  | || ||     |    || |||||  ||||||
Db    129 ACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTACTGGGGATATCCAAGTGAGG 188

Qy    310 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 369
          ||||||||||  ||||||||||| || || ||    ||  | || ||||||   ||
Db    189 AATATGAAGTTGTTACTGAAGATGGCTACATTCTGGGGGTCTATAGAATTCCTTATGGGA 248

Qy    370 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 429
              | || |||| |  ||    || |||||||  |   ||||||||||  |  |||
Db    249 AGAAAAATTCTGAGAATATCGGCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTG 308

Qy    430 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 489
            |  || |  |||||||||  | || ||||| ||||| |||||| |||||||| ||||
Db    309 CATCAGCCACAAACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAG 368

Qy    490 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 549
          ||||||| | ||| |||||| ||||||||||  | || ||  | ||||| || ||| |
Db    369 ATGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCCGGAAAAATG 428

Qy    550 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 609
                ||   ||||  || ||| ||||||||||||||: | ||||||  |||||||  |
Db    429 TATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGGCTAAAT 488

Qy    610 ttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 669
            ||||||||| ||    ||| |||| ||| | |||||||| || || || |||||  | |
Db    489 ATGACCTTCCAGCCACCATAGACTTCATTGTACAGAAAACTGGACAAGAGAAGATACACT 548
```

```
Qy    670 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 729
          |||| ||  | || ||||||||||| || || |||||||| ||||| ||||  || |  |
Db    549 ATGTTGGTCACTCTCAGGGCACCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTC 608

Qy    730 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 789
          ||||| | |||||||| | || || ||| ||||| ||  | || ||||| ||| || ||
Db    609 TGGCTAAAAAAATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAG 668

Qy    790 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 849
          |||| |||   |  ||  |||   | | ||  |  |  |  ||||||   | ||||| |
Db    669 AAAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGATATTTGGTA 728

Qy    850 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 909
          | |||   ||  ||   || | |   ||       |  ||||  ||   |  | ||    |
Db    729 ACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTTGGTACGGAAGTGTGCTCAC 788

Qy    910 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 969
          || | | || ||||   | || || ||      | ||  |  |  |||||||  |||  |
Db    789 GGGAGCTGCTAGATCTTCTCTGCAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGA 848

Qy    970 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 1029
          | ||  | ||  |||| ||      ||| |||   |  || | ||  || |||||||||
Db    849 AAAACTTAAATGTGAGTCGCTTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTA 908

Qy   1030 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 1089
            ||| |  || | ||||||  |||   |  || |||||  | || |  || | | |||
Db    909 CTCAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGCCTATAACT 968

Qy   1090 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1149
          ||||  |||        | ||  ||  | | | |||||||  |||  |||       ||   |
Db    969 GGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAACGCCTCCCTACTATGATG 1028

Qy   1150 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1209
          |   ||  ||||| || || |  ||| ||||||  || || || |||  ||  |  |||
Db   1029 TGTCAGCCATGACCGTGCCAATTGCAGTGTGGAACGGTGGCCATGACATCCTGGCTGATC 1088

Qy   1210 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1269
          |  |||| ||   |||||||||  |  |  |  |||||||  | |||||||||  | ||||
Db   1089 CCCAAGATGTCGCAATGCTGCTTCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTC 1148

Qy   1270 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1329
          |   |  --|||| |||| ||||||||||   |||||| |||||     |  |||||||
Db   1149 TTCCCTACAATCACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATG 1208

Qy   1330 aaatcatccatctgatg 1346
          | ||  |     |||||
Db   1209 AGATAGTTACCATGATG 1225
```

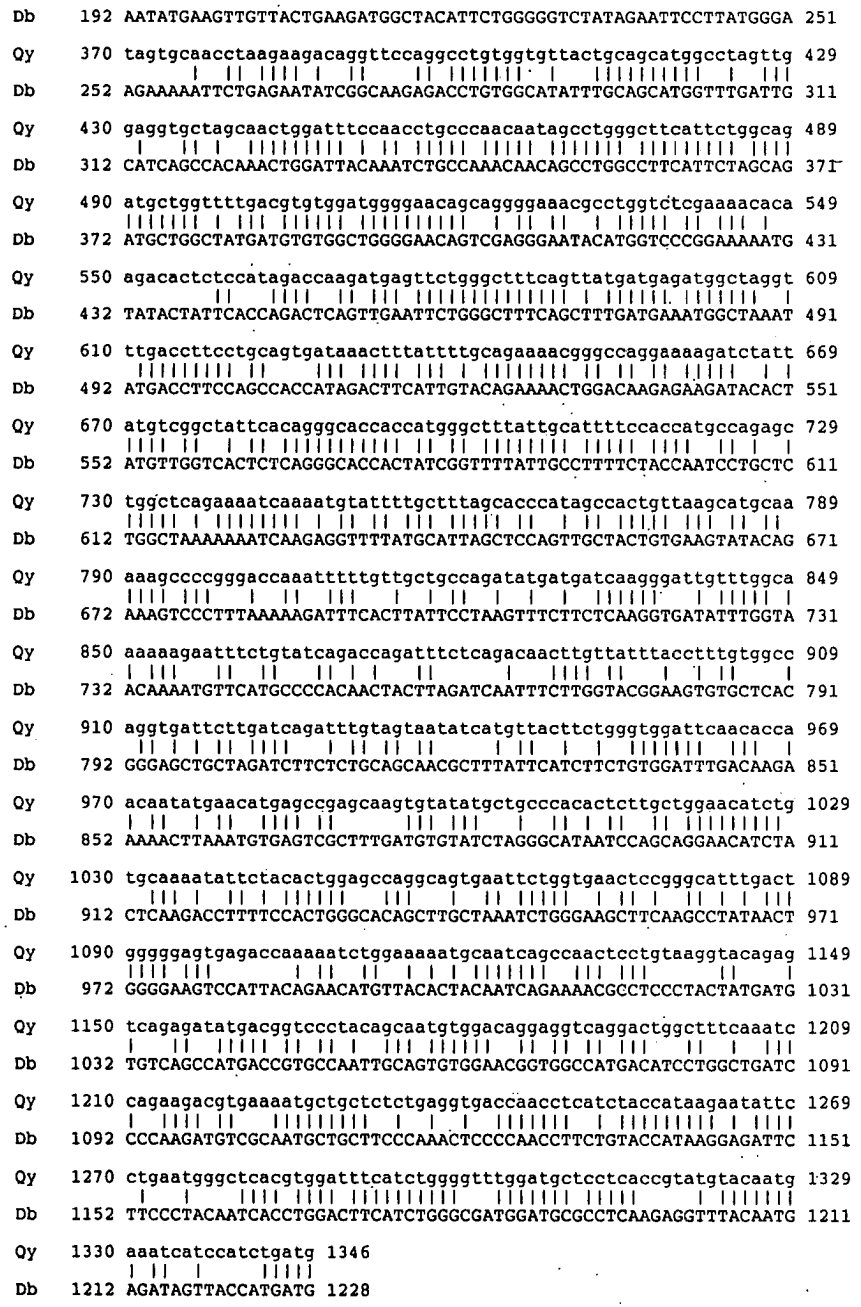
```
RESULT  11
AK009413
LOCUS       AK009413     1345 bp    mRNA            HTC       08-FEB-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310020E22, full insert sequence.
ACCESSION   AK009413
VERSION     AK009413.1  GI:12844193
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310020E22.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods Enzymol. 303, 19-44 (1999)
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Kikuchi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T.,
            Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T.,
            Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Okazaki,Y., Muramatsu,M., Inoue,Y. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1345)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 25.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES                Location/Qualifiers
     source             1. .1345
                        /organism="Mus musculus"
                        /strain="C57BL/6J"
                        /db_xref="taxon:10090"
                        /db_xref="MGD:MGI:1895774"
                        /db_xref="MGD:MGI:1914967"
                        /clone="2310020E22"
                        /sex="male"
                        /tissue_type="tongue"
                        /clone_lib="RIKEN full-length enriched mouse cDNA library"
                        /dev_stage="adult"
     CDS                52. .1239
                        /note="putative"
                        /codon_start=1
                        /protein_id="BAB26272.1"
                        /db_xref="GI:12844194"
                        /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                        SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
```

```
                    FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                    GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
                    FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV
                    YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                    AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                    AED*
BASE COUNT      395 a    288 c    283 g    379 t
ORIGIN


  Query Match           22.4%;  Score 458.6;  DB 192;  Length 1345;
  Best Local Similarity   63.6%;  Pred. No. 2.4e-106;
  Matches  698;  Conservative    0;  Mismatches  399;  Indels    0;  Gaps    0;

Qy    250 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 309
          |||| ||||||  ||||||| ||||| | || ||     |    || |||||  |||||| 
Db    131 ACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTACTGGGGATATCCAAGTGAGG 190

Qy    310 aatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctcgaggcc 369
          |||||||||| |  ||||||||||| || || ||    ||  | || ||||||   ||  
Db    191 AATATGAAGTTGTTACTGAAGATGGCTACATTCTGGGGGTCTATAGAATTCCTTATGGGA 250

Qy    370 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 429
              |  || |||| |  ||    || |||||||  |   |||||||||  |  |||  
Db    251 AGAAAAATTCTGAGAATATCGGCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTG 310

Qy    430 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 489
          |   || |  |||||||| | || ||||| ||||| ||||||| |||||||| |||| 
Db    311 CATCAGCCACAAACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAG 370

Qy    490 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 549
          |||||| | ||| |||||| |||||||||  | || ||  | ||||| || ||| | 
Db    371 ATGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCCGGAAAAATG 430

Qy    550 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 609
                ||   ||||  || ||| |||||||||||||| | |||||| |||||||  |  
Db    431 TATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGGCTAAAT 490

Qy    610 ttgaccttcctgcagtgataaacttttatttttgcagaaaacgggccaggaaaagatctatt 669
          |||||||||  ||    ||| |||| ||| | |||||||| || || || |||||  | | 
Db    491 ATGACCTTCCAGCCACCATAGACTTCATTGTACAGAAAACTGGACAAGAGAAGATACACT 550

Qy    670 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 729
          |||| ||  | || |||||||||||| || || |||||||| ||||| |||| || |  | 
Db    551 ATGTTGGTCACTCTCAGGGCACCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTC 610

Qy    730 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 789
          ||||| | |||||||| | || || ||| ||||| ||  |  || ||||| ||| || || 
Db    611 TGGCTAAAAAAATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAG 670

Qy    790 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 849
          |||| |||   |  ||  |||   |  | ||  |  |  |  ||||||   | ||||| | 
Db    671 AAAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGATATTTGGTA 730

Qy    850 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttaccttttgtggcc 909
          | |||  ||  ||  || | |  ||      |  |||| ||  |  | | ||   | 
Db    731 ACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTTGGTACGGAAGTGTGCTCAC 790

Qy    910 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 969
          || | | || ||||   | || || ||     | ||  |  |  ||||||  |||  | 
Db    791 GGGAGCTGCTAGATCTTCTCTGCAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGA 850

Qy    970 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 1029
          | ||  | ||  |||| ||     ||| |||   |  || | ||  || ||||||||| 
Db    851 AAAACTTAAATGTGAGTCGCTTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTA 910

Qy   1030 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 1089
             ||| |  || | ||||||   |||   |  || |||||  | || |  || | | ||| 
Db    911 CTCAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGCCTATAACT 970

Qy   1090 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1149
          |||| |||       | ||  ||  | | | |||||||  ||| |||      ||   | 
Db    971 GGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAACGCCTCCCTACTATGATG 1030

Qy   1150 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1209
          |   ||  ||||| || || |  ||| ||||||  || || || |||   ||  |  ||| 
Db   1031 TGTCAGCCATGACCGTGCCAATTGCAGTGTGGAACGGTGGCCATGACATCCTGGCTGATC 1090

Qy   1210 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1269
          |  |||| ||   |||||||||  |  |  |  |||||||  | |||||||| | |||| 
Db   1091 CCCAAGATGTCGCAATGCTGCTTCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTC 1150

Qy   1270 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1329
            |   |    |||| |||| ||||||||||   ||||||| |||||     | ||||||| 
Db   1151 TTCCCTACAATCACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATG 1210

Qy   1330 aaatcatccatctgatg 1346
          | ||  |     ||||| 
Db   1211 AGATAGTTACCATGATG 1227


RESULT  12
AK009473
LOCUS       AK009473     1345 bp    mRNA            HTC       08-FEB-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310022H17, full insert sequence.
ACCESSION   AK009473
VERSION     AK009473.1  GI:12844293
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310022H17.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods Enzymol. 303, 19-44 (1999)
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Kikuchi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T.,
            Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T.,
            Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Okazaki,Y., Muramatsu,M., Inoue,Y. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1345)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Inotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
```

```
  JOURNAL    Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
             Physical and Chemical Research (RIKEN), Laboratory for Genome
             Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
             RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
             Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
             URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
             Fax:81-45-503-9216)
  COMMENT    Please visit our web site (http://genome.gsc.riken.go.jp/) for
             further details.
             cDNA library was prepared and sequenced in Mouse Genome
             Encyclopedia Project of Genome Exploration Research Group in Riken
             Genomic Sciences Center and Genome Science Laboratory in RIKEN.
             Division of Experimental Animal Research in Riken contributed to
             prepare mouse tissues. First strand cDNA was primed with a primer
             [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTVN 3'], cDNA was
             prepared by using trehalose thermo-activated reverse transcriptase
             and subsequently enriched for full-length by cap-trapper. cDNA went
             through one round of normalization to Rot = 5.0 and subtraction to
             Rot = 25.0. Second strand cDNA was prepared with the primer adapter
             of sequence [5'
             GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved
             with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
             Host: SOLR.
  FEATURES             Location/Qualifiers
       source          1. .1345
                       /organism="Mus musculus"
                       /strain="C57BL/6J"
                       /db_xref="taxon:10090"
                       /db_xref="MGD:MGI:1898391"
                       /db_xref="MGD:MGI:1914967"
                       /clone="2310022H17"
                       /sex="male"
                       /tissue_type="tongue"
                       /clone_lib="RIKEN full-length enriched mouse cDNA library"
                       /dev_stage="adult"
       CDS             52. .1239
                       /note="putative"
                       /codon_start=1
                       /protein_id="BAB26312.1"
                       /db_xref="GI:12844294"
                       /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                       SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
                       FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                       GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
                       FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV
                       YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                       AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                       AED"
  BASE COUNT      395 a    288 c    283 g    379 t
  ORIGIN


    Query Match           22.4%;  Score 458.6;  DB 192;  Length 1345;
    Best Local Similarity  63.6%;  Pred. No. 2.4e-106;
    Matches  698;  Conservative    0;  Mismatches  399;  Indels    0;  Gaps    0;

Qy    250 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 309
          |||| ||||||  |||||||  |||||  | || ||     |   || |||||  ||||||
Db    131 ACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTACTGGGGATATCCAAGTGAGG 190

Qy    310 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 369
          ||||||||||  |  ||||||||||| || || ||    ||  | || ||||||   ||
Db    191 AATATGAAGTTGTTACTGAAGATGGCTACATTCTGGGGGTCTATAGAATTCCTTATGGGA 250

Qy    370 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 429
              |  || |||| |  ||    || |||||||  |   ||||||||||  |  |||
Db    251 AGAAAAATTCTGAGAATATCGGCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTG 310

Qy    430 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 489
          |   || |  |||||||||| | || |||||  |||||  |||||||  |||||||| ||||
Db    311 CATCAGCCACAAACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAG 370

Qy    490 atgctggtttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 549
          ||||||| | ||| ||||||  |||||||||  | || ||  | |||||  || ||| |
Db    371 ATGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCCGGAAAAATG 430

Qy    550 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 609
              ||   ||||  || ||| ||||||||||||||| | ||||||  |||||||  |
Db    431 TATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGGCTAAAT 490

Qy    610 ttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 669
          |||||||||  ||     ||| |||| ||| | ||||||||  || || || |||||  | |
Db    491 ATGACCTTCCAGCCACCATAGACTTCATTGTACAGAAAACTGGACAAGAGAAGATACACT 550

Qy    670 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 729
          ||||  ||   | ||  |||||||||||  ||  ||  |||||||| ||||| ||||  ||  |  |
Db    551 ATGTTGGTCACTCTCAGGGCACCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTC 610

Qy    730 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 789
          |||||  |  ||||||||  |  ||  ||  |||  |||||  ||   |  ||  |||||  |||  ||  ||
Db    611 TGGCTAAAAAAATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAG 670

Qy    790 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 849
          ||||  |||   |  ||  |||   |  |  ||  |  |  |  ||||||   |  |||||  |
Db    671 AAAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGATATTTGGTA 730

Qy    850 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 909
          |  |||   ||  ||   ||  |  |   ||        |   ||||  ||   |  |  ||    |
Db    731 ACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTTGGTACGGAAGTGTGCTCAC 790

Qy    910 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 969
          ||  |  |  ||  ||||    |  ||  ||  ||      |  ||   |   |  |||||||  |||  |
Db    791 GGGAGCTGCTAGATCTTCTCTGCAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGA 850

Qy    970 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 1029
          |  ||   |  ||  |||  ||      |||  |||    |   ||  |  ||  |||||||||
Db    851 AAAACTTAAATGTGAGTCGCTTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTA 910

Qy   1030 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 1089
            |||  |   ||  |  ||||||    |||    |  ||  |||||   |  ||  |  |  |||
Db    911 CTCAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGCCTATAACT 970

Qy   1090 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1149
          ||||  |||         |  ||   ||  |  |  |||||||   |||  |||       ||    |
Db    971 GGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAACGCCTCCCTACTATGATG 1030

Qy   1150 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1209
          |   ||  ||||| || ||  |  |||  ||||||   || || ||  |||    ||  |  |||
Db   1031 TGTCAGCCATGACCGTGCCAATTGCAGTGTGGAACGGTGGCCATGACATCCTGGCTGATC 1090

Qy   1210 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1269
          |  |||| ||   |||||||||  |  |  |  |||||||  | |||||||||  | ||||
Db   1091 CCCAAGATGTCGCAATGCTGCTTCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTC 1150

Qy   1270 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1329
            |   |    |||| |||| ||||||||||   |||||||  |||||     | |||||||
Db   1151 TTCCCTACAATCACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATG 1210

Qy   1330 aaatcatccatctgatg 1346
          | ||  |     |||||
Db   1211 AGATAGTTACCATGATG 1227


RESULT  13
AK009359
LOCUS       AK009359     1346 bp    mRNA            HTC       08-FEB-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310015H15, full insert sequence.
ACCESSION   AK009359
VERSION     AK009359.1  GI:12844108
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310015H15.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods Enzymol. 303, 19-44 (1999)
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Kikuchi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T.,
            Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T.,
            Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Okazaki,Y., Muramatsu,M., Inoue,Y. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1346)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 25.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES             Location/Qualifiers
     source          1..1346
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /db_xref="MGD:MGI:1895805"
                     /db_xref="MGD:MGI:1914967"
                     /clone="2310015H15"
                     /sex="male"
                     /tissue_type="tongue"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             53..1240
                     /note="putative"
                     /codon_start=1
                     /protein_id="BAB26240.1"
                     /db_xref="GI:12844109"
                     /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                     SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASAKNWITNLPNNSLA
                     FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                     GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKIHFIPK
                     FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV
                     YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                     AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                     AED"
BASE COUNT      396 a    289 c    282 g    379 t
ORIGIN

  Query Match             22.4%;  Score 458.6;  DB 192;  Length 1346;
  Best Local Similarity   63.6%;  Pred. No. 2.4e-106;
  Matches  693;  Conservative    0;  Mismatches  399;  Indels    0;  Gaps    0;

Qy      250 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 309
            |||| ||||||  ||||||| ||||| | || ||     |    || |||||  ||||||
Db      132 ACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTACTGGGGATATCCAAGTGAGG 191

Qy      310 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 369
            |||||||||| |  |||||||||| || || ||    ||  | || ||||||   ||  
Db      192 AATATGAAGTTGTTACTGAAGATGGCTACATTCTGGGGGTCTATAGAATTCCTTATGGGA 251

Qy      370 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 429
                  |  || ||| |  ||    || ||||||||  | ||  ||||||||||  |  |||
Db      252 AGAAAAATTCTGAGAATATCGGCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTG 311

Qy      430 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 489
             |   || |  |||||||| | || |||||| ||||| ||||||||| |||||||| ||||
Db      312 CATCAGCCAAAAACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAG 371

Qy      490 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 549
            |||||||| |||| ||||||| ||||||||||  | || ||  | ||||| || ||| |  
Db      372 ATGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCCGGAAAAATG 431

Qy      550 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 609
                   ||   ||||  || ||| ||||||||||||||| | |||||| |||||||||  |
Db      432 TATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGGCTAAAT 491

Qy      610 ttgaccttcctgcagtgataaactttattttgcagaaaacgggccaggaaaagatctatt 669
             ||||||||| ||    ||| |||| ||| | ||||||| || || || |||||  | |
Db      492 ATGACCTTCCAGCCACCATAGACTTCATTGTACAGAAAACTGGACAAGAGAAGATACACT 551

Qy      670 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 729
            |||| ||  | || ||||||||||| || || |||||||| ||||| ||||  || |  |
Db      552 ATGTTGGTCACTCTCAGGGCACCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTC 611

Qy      730 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 789
            ||||| | |||||||| | || || ||| |||| ||  | || ||||| ||| || || 
Db      612 TGGCTAAAAAAATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAG 671

Qy      790 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 849
            |||| |||    |  ||  ||   ||  | ||  |  |  |  ||||||   | ||||| |
Db      672 AAAGTCCCTTTAAAAAGATTCACTTTATTCCTAAGTTTCTTCTCAAGGTGATATTTGGTA 731

Qy      850 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 909
            | |||   ||  ||   || | |   ||      |   |||| ||   |  | ||    |
Db      732 ACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTTGGTACGGAAGTGTGCTCAC 791

Qy      910 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 969
             || | | || ||||   | || || ||     | ||  |  |  |||||||  |||  |
Db      792 GGGAGCTGCTAGATCTTCTCTGCAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGA 851
```

```
Qy    970 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 1029
          | ||  | ||  |||| ||      ||| |||   |  || | ||  || ||||||||||
Db    852 AAAACTTAAATGTGAGTCGCTTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTA 911

Qy   1030 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 1089
            ||| |  || | ||||||   |||   |  || |||||  || |  ||  | | |||
Db    912 CTCAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGCCTATAACT 971

Qy   1090 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaagctacagag 1149
          |||| |||       | ||  ||  | | | |||||||  ||| |||      ||    |
Db    972 GGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAACGCCTCCCTACTATGATG 1031

Qy   1150 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1209
          |   ||  ||||| || || |  ||| ||||||  || || || |||   ||  |  |||
Db   1032 TGTCAGCCATGACCGTGCCAATTGCAGTGTGGAACGGTGGCCATGACATCCTGGCTGATC 1091

Qy   1210 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1269
          |  |||| ||   |||||||||  |  | |  |||||||  | |||||||| | ||||
Db   1092 CCCAAGATGTCGCAATGCTGCTTCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTC 1151

Qy   1270 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1329
            |   |   |||| |||| ||||||||||   |||||| |||||     | |||||||
Db   1152 TTCCCTACAATCACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATG 1211

Qy   1330 aaatcatccatctgatg 1346
          | ||  |     |||||
Db   1212 AGATAGTTACCATGATG 1228


RESULT  14
AK009523
LOCUS       AK009523     1346 bp    mRNA            HTC       08-FEB-2001
DEFINITION  Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310028I09, full insert sequence.
ACCESSION   AK009523
VERSION     AK009523.1  GI:12844369
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2310028I09.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods Enzymol. 303, 19-44 (1999)
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Kikuchi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T.,
            Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T.,
            Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Okazaki,Y., Muramatsu,M., Inoue,Y. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1346)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 25.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES              Location/Qualifiers
     source           1. .1346
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /db_xref="MGD:MGI:1895759"
                      /db_xref="MGD:MGI:1914967"
                      /clone="2310028I09"
                      /sex="male"
                      /tissue_type="tongue"
                      /clone_lib="RIKEN full-length enriched mouse cDNA library"
                      /dev_stage="adult"
     CDS              53. .1240
                      /note="putative"
                      /codon_start=1
                      /protein_id="BAB26338.1"
                      /db_xref="GI:12844370"
                      /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                      SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
                      FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEFWAFSFDEMAKYDLPATIDFIVQKT
                      GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
                      FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV
                      YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                      AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                      AED"
BASE COUNT      395 a    289 c    283 g    379 t
ORIGIN


  Query Match             22.4%;  Score 458.6;  DB 192;  Length 1346;
  Best Local Similarity   63.6%;  Pred. No. 2.4e-106;
  Matches  698;  Conservative    0;  Mismatches  399;  Indels    0;  Gaps    0;

Qy    250 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 309
          |||| ||||||  ||||||| ||||| | || ||     |    || |||||  ||||||
Db    132 ACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTACTGGGGATATCCAAGTGAGG 191

Qy    310 aatatgaagtcgcaactgaagatgggtatatccctttctgttaacaggattcctcgaggcc 369
          |||||||||| |  |||||||||| || || ||    ||  | || ||||||   ||
```



RESULT 15
AK009729

LOCUS AK009729 1346 bp mRNA HTC 08-FEB-2001
DEFINITION Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310040L03, full insert sequence.
ACCESSION AK009729
VERSION AK009729.1 GI:12844702
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library clone:2310040L03.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (sites)
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Methods Enzymol. 303, 19-44 (1999)
REFERENCE 2 (sites)
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
REFERENCE 3 (sites)
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Kikuchi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Okazaki,Y., Muramatsu,M., Inoue,Y. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
REFERENCE 4 (sites)
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5 (bases 1 to 1346)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Gemomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 25.0. Second strand cDNA was prepared with the primer adapter of sequence [5'

```
          GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved
          with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
          Host: SOLR.
FEATURES             Location/Qualifiers
     source          1. .1346
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /db_xref="MGD:MGI:1898390"
                     /db_xref="MGD:MGI:1914967"
                     /clone="2310040L03"
                     /sex="male"
                     /tissue_type="tongue"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             53. .1240
                     /note="putative"
                     /codon_start=1
                     /protein_id="BAB26466.1"
                     /db_xref="GI:12844703"
                     /translation="MWLLLVTSVLSAFGGAHGLFGKLGPKNPEANMNVSQMITYWGYP
                     SEEYEVVTEDGYILGVYRIPYGKKNSENIGKRPVAYLQHGLIASATNWITNLPNNSLA
                     FILADAGYDVWLGNSRGNTWSRKNVYYSPDSVEEFWAFSFDEMAKYDLPATIDFIVQKT
                     GQEKIHYVGHSQGTTIGFIAFSTNPALAKKIKRFYALAPVATVKYTESPFKKISLIPK
                     FLLKVIFGNKMFMPHNYLDQFLGTEVCSRELLDLLCSNALFIFCGFDKKNLNVSRFDV
                     YLGHNPAGTSTQDLFHWAQLAKSGKLQAYNWGSPLQNMLHYNQKTPPYYDVSAMTVPI
                     AVWNGGHDILADPQDVAMLLPKLPNLLYHKEILPYNHLDFIWAMDAPQEVYNEIVTMM
                     AED"
BASE COUNT      395 a    288 c    283 g    380 t
ORIGIN


  Query Match           22.4%;  Score 458.6;  DB 192;  Length 1346;
  Best Local Similarity  63.6%;  Pred. No. 2.4e-106;
  Matches  698;  Conservative    0;  Mismatches  399;  Indels    0;  Gaps    0;

Qy    250 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 309
          |||| ||||||  ||||||| ||||| | || ||     |    || |||||  ||||||
Db    132 ACCCTGAAGCAAACATGAATGTTAGTCAGATGATAACTTACTGGGGATATCCAAGTGAGG 191

Qy    310 aatatgaagtcgcaactgaagatgggtatatccttttctgttaacaggattcctcgaggcc 369
          ||||||||| |  |||||||||| || || ||    ||  | || ||||||   ||
Db    192 AATATGAAGTTGTTACTGAAGATGGCTACATTCTGGGGGTCTATAGAATTCCTTATGGGA 251

Qy    370 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 429
              |  || |||| |  ||    || |||||||  |   ||||||||||  |  |||
Db    252 AGAAAAATTCTGAGAATATCGGCAAGAGACCTGTGGCATATTTGCAGCATGGTTTGATTG 311

Qy    430 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 489
           |   || |  |||||||||  | || ||||| ||||| ||||||| |||||||||  ||||
Db    312 CATCAGCCACAAACTGGATTACAAATCTGCCAAACAACAGCCTGGCCTTCATTCTAGCAG 371

Qy    490 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 549
          ||||||| | ||| |||||| ||||||||||  | || ||  | ||||| || ||| |
Db    372 ATGCTGGCTATGATGTGTGGCTGGGGAACAGTCGAGGGAATACATGGTCCCGGAAAAATG 431

Qy    550 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 609
                ||   ||||  || ||| |||||||||||||| | |||||| |||||||  |
Db    432 TATACTATTCACCAGACTCAGTTGAATTCTGGGCTTTCAGCTTTGATGAAATGGCTAAAT 491

Qy    610 ttgaccttcctgcagtgataaactttatttttgcagaaaacgggccaggaaaagatctatt 669
           |||||||||  ||    ||| |||| ||| | ||||||||  || || || |||||  | |
Db    492 ATGACCTTCCAGCCACCATAGACTTCATTGTACAGAAAACTGGACAAGAGAAGATACACT 551

Qy    670 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 729
          |||| ||  | || ||||||||||| || || |||||||| ||||| ||||  || |  |
Db    552 ATGTTGGTCACTCTCAGGGCACCACTATCGGTTTTATTGCCTTTTCTACCAATCCTGCTC 611

Qy    730 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 789
          |||||  | |||||||| | || || ||| ||||| ||. | || |||||.||| || ||
Db    612 TGGCTAAAAAAATCAAGAGGTTTTATGCATTAGCTCCAGTTGCTACTGTGAAGTATACAG 671

Qy    790 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 849
          |||| |||   |  ||  |||   |  | ||  |  |  |  ||||||   | |||||  |
Db    672 AAAGTCCCTTTAAAAAGATTTCACTTATTCCTAAGTTTCTTCTCAAGGTGATATTTGGTA 731

Qy    850 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 909
          | |||   ||  ||   || | |   ||      |   |||| ||   |  | ||    |
Db    732 ACAAAATGTTCATGCCCCACAACTACTTAGATCAATTTCTTGGTACGGAAGTGTGCTCAC 791

Qy    910 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 969
          || | | || |||   | || || ||     | ||  |  |  |||||||  |||  |
Db    792 GGGAGCTGCTAGATCTTCTCTGCAGCAACGCTTTATTCATCTTCTGTGGATTTGACAAGA 851

Qy    970 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 1029
          | ||  | ||  |||| ||      |||.|||   |  || | ||  || |||||||||
Db    852 AAAACTTAAATGTGAGTCGCTTTGATGTGTATCTAGGGCATAATCCAGCAGGAACATCTA 911

Qy   1030 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 1089
            ||| |  || | ||||||   |||   |  || |||||  | || |  || | | |||
Db    912 CTCAAGACCTTTTCCACTGGGCACAGCTTGCTAAATCTGGGAAGCTTCAAGCCTATAACT 971

Qy   1090 ggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1149
          |||| |||       | ||  ||  | | | |||||||  ||| |||      ||   |
Db    972 GGGGAAGTCCATTACAGAACATGTTACACTACAATCAGAAAACGCCTCCCTACTATGATG 1031

Qy   1150 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1209
          |   |  ||||| || || |  ||| ||||||  || || || |||   ||  |  |||
Db   1032 TGTCAGCCATGACCGTGCCAATTGCAGTGTGGAACGGTGGCCATGACATCCTGGCTGATC 1091

Qy   1210 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1269
          |  |||| ||  |||||||||  |  |  |  |||||||  | |||||||||  | ||||
Db   1092 CCCAAGATGTCGCAATGCTGCTTCCCAAACTCCCCAACCTTCTGTACCATAAGGAGATTC 1151

Qy   1270 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1329
          |   |   |||| |||| ||||||||||  |||||||  |||||     | |||||||
Db   1152 TTCCCTACAATCACCTGGACTTCATCTGGGCGATGGATGCGCCTCAAGAGGTTTACAATG 1211

Qy   1330 aaatcatccatctgatg 1346
          | ||  |     |||||
Db   1212 AGATAGTTACCATGATG 1228


Search completed: August 24, 2001, 21:22:40
Job time: 5547 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 24, 2001, 22:38:56 ; Search time 97.86 Seconds
(without alignments)
2454.893 Million cell updates/sec

Title: US-09-333-159-46
Perfect score: 1269
Sequence: 1 atgttggaaaccttgtcaag.....gacggtgtgaggccgtattg 1269

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	444.2	35.0	1137	1	US-08-227-108-2	Sequence 2, Appli
	2	444.2	35.0	1137	2	US-09-073-674-2	Sequence 2, Appli
	3	444.2	35.0	1140	1	US-08-227-108-4	Sequence 4, Appli
	4	444.2	35.0	1140	2	US-09-073-674-4	Sequence 4, Appli
	5	444.2	35.0	1146	1	US-08-227-108-6	Sequence 6, Appli
	6	444.2	35.0	1146	2	US-09-073-674-6	Sequence 6, Appli
	7	444.2	35.0	1528	1	US-08-227-108-1	Sequence 1, Appli
	8	444.2	35.0	1528	2	US-09-073-674-1	Sequence 1, Appli
	9	48.2	3.8	178	1	US-08-751-782-2	Sequence 2, Appli
	10	48.2	3.8	178	2	US-08-925-171-2	Sequence 2, Appli
	11	46.4	3.7	43676	3	US-09-356-952-12	Sequence 12, Appl
	12	36.6	2.9	3271	2	US-08-852-806-1	Sequence 1, Appli
	13	36.6	2.9	3271	3	US-09-163-669-1	Sequence 1, Appli
	14	36.6	2.9	3282	1	US-08-276-852-154	Sequence 154, App
c	15	36.6	2.9	3282	1	US-08-276-852-169	Sequence 169, App
	16	36.6	2.9	3282	1	US-08-899-575-154	Sequence 154, App
c	17	36.6	2.9	3282	1	US-08-899-575-169	Sequence 169, App
	18	36.6	2.9	3282	1	US-08-899-575-154	Sequence 154, App
c	19	36.6	2.9	3282	1	US-08-899-575-169	Sequence 169, App
	20	36.6	2.9	3282	5	PCT-US95-08743-154	Sequence 154, App
c	21	36.6	2.9	3282	5	PCT-US95-08743-169	Sequence 169, App
	22	36.6	2.9	13254	1	US-08-276-852-156	Sequence 156, App
c	23	36.6	2.9	13254	1	US-08-276-852-170	Sequence 170, App
	24	36.6	2.9	13254	1	US-08-899-575-156	Sequence 156, App
c	25	36.6	2.9	13254	1	US-08-899-575-170	Sequence 170, App
	26	36.6	2.9	13254	1	US-08-899-575-156	Sequence 156, App
c	27	36.6	2.9	13254	1	US-08-899-575-170	Sequence 170, App
	28	36.6	2.9	13254	5	PCT-US95-08743-156	Sequence 156, App
c	29	36.6	2.9	13254	5	PCT-US95-08743-170	Sequence 170, App
	30	33.6	2.6	8392	1	US-08-080-255-6	Sequence 6, Appli
	31	33.6	2.6	8392	3	US-08-465-713-6	Sequence 6, Appli
	32	33.6	2.6	8392	5	PCT-US93-05857-6	Sequence 6, Appli
c	33	32.6	2.6	7218	1	US-08-232-463-14	Sequence 14, Appl
c	34	32	2.5	2197	1	US-08-233-005-3	Sequence 3, Appli
c	35	32	2.5	2197	1	US-08-428-943-3	Sequence 3, Appli
c	36	32	2.5	2197	3	US-09-016-649-3	Sequence 3, Appli
c	37	32	2.5	2197	5	PCT-US95-04858-3	Sequence 3, Appli
	38	32	2.5	8342	3	US-08-545-860D-63	Sequence 63, Appl
	39	32	2.5	8342	5	PCT-US94-04496-63	Sequence 63, Appl
	40	31	2.4	3212	3	US-08-673-814-1	Sequence 1, Appli
c	41	30.4	2.4	1529	1	US-08-336-778-1	Sequence 1, Appli
c	42	30.2	2.4	4982	3	US-08-699-103B-1	Sequence 1, Appli
	43	30.2	2.4	10079	2	US-08-476-866-20	Sequence 20, Appl
c	44	30	2.4	3639	2	US-08-737-524B-26	Sequence 26, Appl
	45	30	2.4	5178	2	US-08-474-169-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-227-108-2
; Sequence 2, Application US/08227108
; Patent No. 5807726
; GENERAL INFORMATION:
; APPLICANT: Blanchard, Claire
; APPLICANT: Benicourt, Claude
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: Recombinant Dog Gastric Lipase
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/227,108
; FILING DATE: 03-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fanucci, Allan A.
; REGISTRATION NUMBER: 30,256
; REFERENCE/DOCKET NUMBER: 7620-033
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1137 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1137
US-08-227-108-2

Query Match 35.0%; Score 444.2; DB 1; Length 1137;
Best Local Similarity 62.8%; Pred. No. 5.6e-139;
Matches 689; Conservative 0; Mismatches 408; Indels 0; Gaps 0;

```
Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
Db     26 ACCCTGAAGTGACCATGAATATAAGTCAGATGATCACCTACTGGGGATACCCAGCTGAGG 85

Qy    185 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 244
Db     86 AATATGAAGTTGTGACCGAAGACGGTTATATCCTTGGGATCGACAGAATTCCTTATGGGA 145

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
Db    146 GGAAAAATTCAGAGAATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCG 205

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
Db    206 CATCAGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCCTGGCCG 265

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
Db    266 ACGCCGGGTACGACGTGTGGCTGGGGAACAGCAGGGGCAACACCTGGGCCAGGAGGAATC 325

Qy    425 agacactctccatagaccaagatgagttctgggctttcagttatgatgagatggctaggt 484
Db    326 TGTACTACTCGCCCGACTCCGTCGAATTCTGGGCTTTCAGCTTTGACGAGATGGCTAAAT 385

Qy    485 ttgaccttcctgcagtgataaactttatttttgcagaaaacgggccaggaaaagatctatt 544
Db    386 ATGACCTTCCCGCCACCATTGACTTCATCTTGAAGAAAACGGGACAGGACAAGCTACACT 445

Qy    545 atgtcggctattcacagggcaccaccatgggctttattgcattttccaccatgccagagc 604
Db    446 ACGTTGGCCATTCCCAGGGCACCACCATTGGTTTCATCGCCTTTTCCACCAATCCCAAGC 505

Qy    605 tggctcagaaaatcaaaatgtattttgctttagcacccatagccactgttaagcatgcaa 664
Db    506 TGGCGAAACGGATCAAAACCTTCTATGCATTAGCTCCCGTTGCCACCGTGAAGTACACCG 565

Qy    665 aaagccccgggaccaaatttttgttgctgccagatatgatgatcaagggattgtttggca 724
Db    566 AAACCCTGTTAAACAAACTCATGCTCGTCCCTTCGTTCCTCTTCAAGCTTATATTTGGAA 625

Qy    725 aaaaagaatttctgtatcagaccagatttctcagacaacttgttatttacctttgtggcc 784
Db    626 ACAAAATATTCTACCCACACCACTTCTTTGATCAATTTCTCGCCACCGAGGTATGCTCCC 685

Qy    785 aggtgattcttgatcagatttgtagtaatatcatgttacttctgggtggattcaacacca 844
Db    686 GCGAGACGGTGGATCTCCTCTGCAGCAACGCCCTGTTTATCATTTGTGGATTTGACACTA 745

Qy    845 acaatatgaacatgagccgagcaagtgtatatgctgcccacactcttgctggaacatctg 904
Db    746 TGAACTTGAACATGAGTCGCTTGGATGTGTATCTGTCACATAATCCAGCAGGAACATCGG 805

Qy    905 tgcaaaatattctacactggagccaggcagtgaattctggtgaactccgggcatttgact 964
Db    806 TTCAGAACGTGCTCCACTGGTCCCAGGCTGTTAAGTCTGGGAAGTTCCAAGCTTTTGACT 865

Qy    965 gggggagtgagaccaaaaatctggaaaaatgcaatcagccaactcctgtaaggtacagag 1024
Db    866 GGGGAAGCCCAGTTCAGAACATGATGCACTATCATCAGAGCATGCCTCCCTACTACAACC 925

Qy   1025 tcagagatatgacggtccctacagcaatgtggacaggaggtcaggactggctttcaaatc 1084
Db    926 TGACAGACATGCATGTGCCAATCGCAGTGTGGAACGGTGGCAACGACTTGCTGGCCGACC 985

Qy   1085 cagaagacgtgaaaatgctgctctctgaggtgaccaacctcatctaccataagaatattc 1144
Db    986 CTCACGATGTTGACCTTTTGCTTTCCAAGCTCCCCAATCTCATTTACCACAGGAAGATTC 1045

Qy   1145 ctgaatgggctcacgtggatttcatctggggtttggatgctcctcaccgtatgtacaatg 1204
Db   1046 CTCCTTACAATCACTTGGACTTTATCTGGGCCATGGATGCCCCTCAAGCGGTTTACAATG 1105

Qy   1205 aaatcatccatctgatg 1221
Db   1106 AAATTGTTTCCATGATG 1122
```

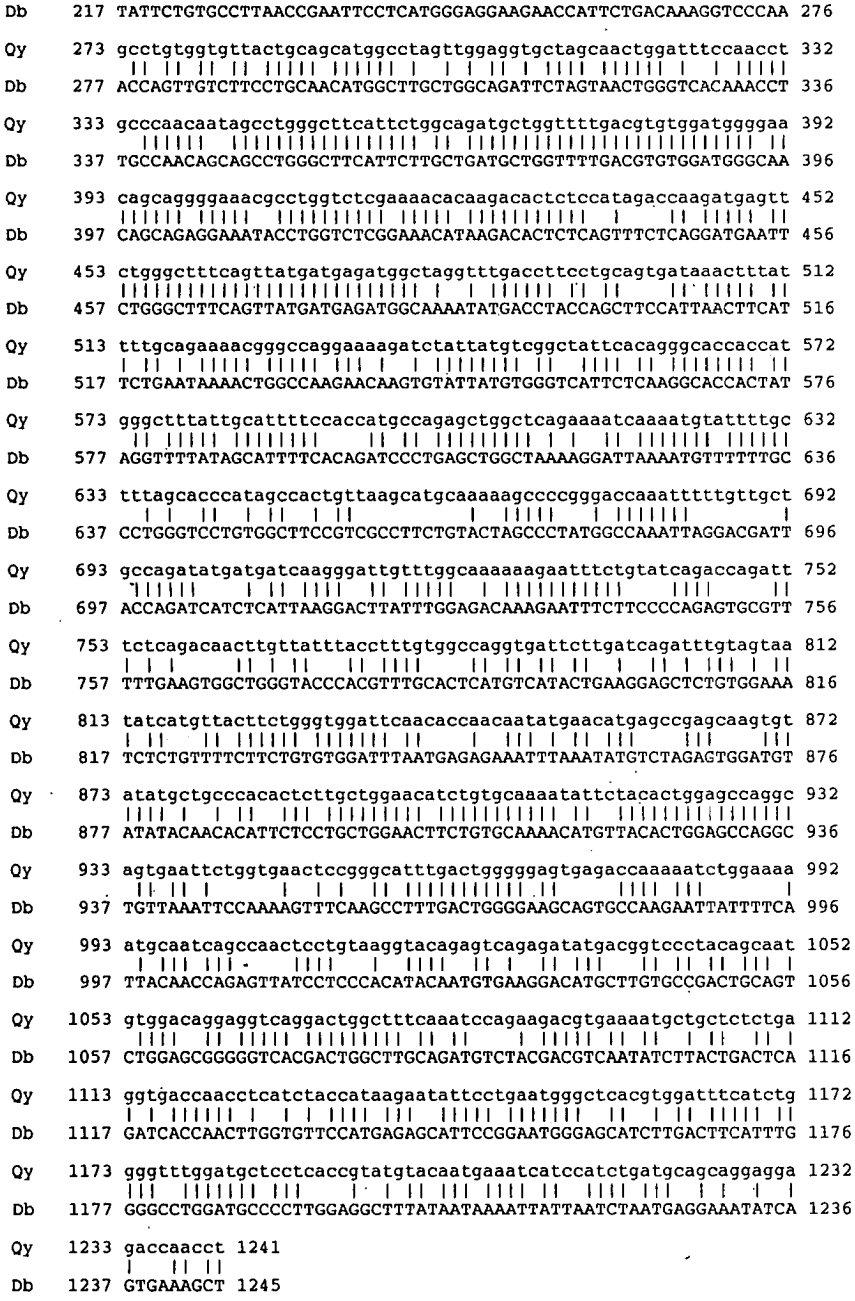
```
RESULT   2
US-09-073-674-2
; Sequence 2, Application US/09073674
; Patent No. 5998189
;  GENERAL INFORMATION:
;    APPLICANT:  Blanchard, Claire
;    APPLICANT:  Benicourt, Claude
;    APPLICANT:  Junien, Jean-Louis
;    TITLE OF INVENTION:  Recombinant Dog Gastric Lipase
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Warner-Lambert Company
;      STREET:  2800 Plymouth Road
;      CITY:  Ann Arbor
;      STATE:  Michigan
;      COUNTRY:  U.S.A.
;      ZIP:  48105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/073,674
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crissey, Todd M.
;      REGISTRATION NUMBER:  37,807
;      REFERENCE/DOCKET NUMBER:  5072-D1-66-TMC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  734 622-7530
;      TELEFAX:  734 622-1553
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1137 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1137
US-09-073-674-2
```

```
  Query Match             35.0%;  Score 444.2;  DB 2;  Length 1137;
  Best Local Similarity   62.8%;  Pred. No. 5.6e-139;
  Matches  689;  Conservative    0;  Mismatches  408;  Indels    0;  Gaps    0;

Qy    125 acccagaagcattcatgaatattagtgaaatcatccaacatcaaggctatccctgtgagg 184
Db     26 ACCCTGAAGTGACCATGAATATAAGTCAGATGATCACCTACTGGGGATACCCAGCTGAGG 85

Qy    185 aatatgaagtcgcaactgaagatgggtatatcctttctgttaacaggattcctcgaggcc 244
Db     86 AATATGAAGTTGTGACCGAAGACGGTTATATCCTTGGGATCGACAGAATTCCTTATGGGA 145

Qy    245 tagtgcaacctaagaagacaggttccaggcctgtggtgttactgcagcatggcctagttg 304
Db    146 GGAAAAATTCAGAGAATATAGGCCGGAGACCTGTTGCATTTTTGCAACACGGTTTGCTCG 205

Qy    305 gaggtgctagcaactggatttccaacctgcccaacaatagcctgggcttcattctggcag 364
Db    206 CATCAGCCACAAACTGGATCTCCAACCTGCCCAACAACAGCCTGGCCTTCATCCTGGCCG 265

Qy    365 atgctggttttgacgtgtggatggggaacagcaggggaaacgcctggtctcgaaaacaca 424
```

```
Qy     393 cagcaggggaaacgcctggtctcgaaaacacaagacactctccatagaccaagatgagtt 452
           |||||| |||||  |||||||||| ||||| ||||||||||  |    || ||||| ||
Db     502 CAGCAGAGGAAATACCTGGTCTCGGAAACATAAGACACTCTCAGTTTCTCAGGATGAATT 561

Qy     453 ctgggctttcagttatgatgagatggctaggtttgaccttcctgcagtgataaaactttat 512
           ||||||||||||||||||||||||||| |  | |||||| || ||    || ||||| ||
Db     562 CTGGGCTTTCAGTTATGATGAGATGGCAAAATATGACCTACCAGCTTCCATTAACTTCAT 621

Qy     513 tttgcagaaaacgggccaggaaaagatctattatgtcggctattcacagggcaccaccat 572
           | || | ||||| ||||| ||| |  | |||||||| ||  |||| || |||||||| ||
Db     622 TCTGAATAAAACTGGCCAAGAACAAGTGTATTATGTGGGTCATTCTCAAGGCACCACTAT 681

Qy     573 gggctttattgcattttccaccatgccagagctggctcagaaaatcaaaatgtattttgc 632
           || ||||| ||||||||    || || |||||||||  | |  || |||||||  ||||||
Db     682 AGGTTTTATAGCATTTTCACAGATCCCTGAGCTGGCTAAAAGGATTAAAATGTTTTTTGC 741

Qy     633 tttagcacccatagccactgttaagcatgcaaaaagccccgggaccaaatttttgttgct 692
           | |  ||  | ||  | ||          |  |||||   | ||||||         |
Db     742 CCTGGGTCCTGTGGCTTCCGTCGCCTTCTGTACTAGCCCTATGGCCAAATTAGGACGATT 801

Qy     693 gccagatatgatgatcaagggattgtttggcaaaaaagaatttctgtatcagaccagatt 752
           ||||||    | || ||||  || |||||  | ||||||||||     ||||     ||
Db     802 ACCAGATCATCTCATTAAGGACTTATTTGGAGACAAAGAATTTCTTCCCCAGAGTGCGTT 861

Qy     753 tctcagacaacttgttatttacctttgtggccaggtgattcttgatcagatttgtagtaa 812
           | | |     || | ||   || ||||    || || || ||  |  || | ||| | ||
Db     862 TTTGAAGTGGCTGGGTACCCACGTTTGCACTCATGTCATACTGAAGGAGCTCTGTGGAAA 921

Qy     813 tatcatgttacttctgggtggattcaacaccaacaatatgaacatgagccgagcaagtgt 872
           |. ||   || |||||| ||||||| ||    |  ||| | || |||    |||    |||
Db     922 TCTCTGTTTTCTTCTGTGTGGATTTAATGAGAGAAATTTAAATATGTCTAGAGTGGATGT 981

Qy     873 atatgctgcccacactcttgctggaacatctgtgcaaaatattctacactggagccaggc 932
           ||||  |  | ||  ||| |||||||||| |||||||||| ||  ||||||||||||||
Db     982 ATATACAACACATTCTCCTGCTGGAACTTCTGTGCAAAACATGTTACACTGGAGCCAGGC 1041

Qy     933 agtgaattctggtgaactccgggcatttgactgggggagtgagaccaaaaatctggaaaa 992
           || || |      |  | |  || |||||||||| ||     |||| |||       |
Db    1042 TGTTAAATTCCAAAAGTTTCAAGCCTTTGACTGGGGAAGCAGTGCCAAGAATTATTTTCA 1101

Qy     993 atgcaatcagccaactcctgtaaggtacagagtcagagatatgacggtccctacagcaat 1052
           | ||| |||     ||||   | ||||  || |  || |||   || || || ||| |
Db    1102 TTACAACCAGAGTTATCCTCCCACATACAATGTGAAGGACATGCTTGTGCCGACTGCAGT 1161

Qy    1053 gtggacaggaggtcaggactggctttcaaatccagaagacgtgaaaatgctgctctctga 1112
           ||||  || ||||| |||||||| || ||    | ||||| || ||  | ||  || |
Db    1162 CTGGAGCGGGGGTCACGACTGGCTTGCAGATGTCTACGACGTCAATATCTTACTGACTCA 1221

Qy    1113 ggtgaccaacctcatctaccataagaatattcctgaatgggctcacgtggatttcatctg 1172
           | | |||||| |  | | |||| |||  ||||| |||||||  ||  | || ||||| ||
Db    1222 GATCACCAACTTGGTGTTCCATGAGAGCATTCCGGAATGGGAGCATCTTGACTTCATTTG 1281

Qy    1173 gggtttggatgctcctcaccgtatgtacaatgaaatcatccatctgatgcagcaggagga 1232
           |||  ||||||| |||    |  | || ||| |||| ||  |||| |||  | |  |  |
Db    1282 GGGCCTGGATGCCCCTTGGAGGCTTTATAATAAAATTATTAATCTAATGAGGAAATATCA 1341

Qy    1233 gaccaacct 1241
           |   || ||
Db    1342 GTGAAAGCT 1350
```

```
RESULT   4
HUMLIPCHL
LOCUS       HUMLIPCHL    2493 bp    mRNA            PRI       07-JAN-1995
DEFINITION  Human lysosomal acid lipase/cholesteryl esterase mRNA, complete
            cds.
ACCESSION   M74775
VERSION     M74775.1  GI:187151
KEYWORDS    lysosomal acid lipase/cholesteryl esterase.
SOURCE      Homo sapiens fibroblast cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2493)
  AUTHORS   Anderson,R.A. and Sando,G.N.
  TITLE     Cloning and expression of cDNA encoding human lysosomal acid
            lipase/cholesteryl ester hydrolase. Similarities to gastric and
            lingual lipases
  JOURNAL   J. Biol. Chem. 266 (33), 22479-22484 (1991)
  MEDLINE   92012192
FEATURES             Location/Qualifiers
     source          1. .2493
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="fibroblast"
                     /map="10"
     mRNA            <1. .2493
                     /gene="LIPA"
                     /note="G00-120-153"
     gene            1. .2493
                     /gene="LIPA"
     CDS             41. .1240
                     /gene="LIPA"
                     /EC_number="3.1.1.13"
                     /note="alternate initiator ATG may be at position 47"
                     /codon_start=1
                     /db_xref="GDB:G00-120-153"
                     /product="lysosomal acid lipase/cholesteryl esterase"
                     /protein_id="AAA59519.1"
                     /db_xref="GI:187152"
                     /translation="MKMRFLGLVVCLVLWPLHSEGSGGKLTAVDPETNMNVSEIISYW
                     GFPSEEYLVETEDGYILCLNRIPHGRKNHSDKGPKPVVFLQHGLLADSSNWVTNLANS
                     SLGFILADAGFDVWMGNSRGNTWSRKHKTLSVSQDEFWAFSYDEMAKYDLPASINFIL
                     NKTGQEQVYYVGHSQGTTIGFIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGR
                     LPDHLIKDLFGDKEFLPQSAFLKWLGTHVCTHVILKELCGNLCFLLCGFNERNLNMSR
                     VDVYTTHSPAGTSVQNMLHWSQAVKFQKFQAFDWGSSAKNYFHYNQSYPPTYNVKDML
                     VPTAVWSGGHDWLADVYDVNILLTQITNLVFHESIPEWEHLDFIWGLDAPWRLYNKII
                     NLMRKYQ"
     sig_peptide     41. .103
                     /gene="LIPA"
                     /note="alternate signal peptide may begin at position 47;
                     as predicted from von Heijne, G. (1986) Nucl. Acids Res.
                     14,4683-4690; G00-120-153"
     mat_peptide     104. .1237
                     /gene="LIPA"
                     /EC_number="3.1.1.13"
                     /note="G00-120-153"
                     /product="lysosomal acid lipase/cholesteryl esterase"
     polyA_signal    2466. .2471
                     /gene="LIPA"
                     /note="G00-120-153"
     polyA_site      2493
                     /gene="LIPA"
                     /note="G00-120-153"
BASE COUNT      692 a    484 c    516 g    801 t
ORIGIN
```

```
  Query Match             41.5%;  Score 526.2;  DB 97;  Length 2493;
  Best Local Similarity   65.3%;  Pred. No. 1.4e-145;
  Matches  790;  Conservative    0;  Mismatches  413;  Indels    6;  Gaps    1;

Qy      39 cagaatggaaatgtggcttctgattctggtggcgtatatgttccagagaaatgtgaattc 98
           ||||||| ||||| || |  ||    ||||||  | | || | |   |      || ||||
Db      37 CAGAATGAAAATGCGGTTCTTGGGGTTGGTGGTCTGTTTGGTTCTCTGGCCCCTGCATTC 96

Qy      99 ag------tacatatgccaactaaagctgtggacccagaagcattcatgaatattagtga 152
           |       |  |   | ||||||||| || ||| ||  ||||||| | |||||
Db      97 TGAGGGGTCTGGAGGGAAACTGACAGCTGTGGATCCTGAAACAAACATGAATGTGAGTGA 156

Qy     153 aatcatccaacatcaaggctatccctgtgaggaatatgaagtcgcaactgaagatgggta 212
           ||| |||    |    || |  ||  |||||||||   ||| |  || ||||||| ||
Db     157 AATTATCTCTTACTGGGGATTCCCTAGTGAGGAATACCTAGTTGAGACAGAAGATGGATA 216

Qy     213 tatccttttctgttaacaggattcctcgaggcctagtgcaacctaagaagacaggttccag 272
           ||| || |   ||||| | |||||||  ||      | | | |    | | |||| |||
```



RESULT 5
A26689

LOCUS A26689 1378 bp DNA PAT 05-APR-1995
DEFINITION Precursor of rabit gastric lipase coding sequence.
ACCESSION A26689
VERSION A26689.1 GI:905029
KEYWORDS .
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequence.
REFERENCE 1 (bases 1 to 1378)
AUTHORS Benicourt,C., Blanchard,C. and Junien,J.L.
TITLE Recombinant gastric lipase from rabbit and pharmaceutical compositions
JOURNAL Patent: EP 0542629-A 9 19-MAY-1993;
INSTITUT DE RECHERCHE JOUVEINAL
FEATURES Location/Qualifiers
source 1. .1378
/organism="synthetic construct"
/db_xref="taxon:32630"
BASE COUNT 397 a 287 c 276 g 418 t
ORIGIN

Query Match 38.6%; Score 489.2; DB 9; Length 1378;
Best Local Similarity 65.0%; Pred. No. 1.4e-134;
Matches 722; Conservative 0; Mismatches 388; Indels 0; Gaps 0;